Analysis of Multivariate Sensor Data for Monitoring of Cultivations

Petersen, Nanna

Publication date:
2010

DTU Chemical Engineering
Department of Chemical and Biochemical Engineering

Nanna Petersen Rønnest

Analysis of Multivariate Sensor Data for Monitoring of Cultivations

PhD Thesis, July 2010

Nanna Petersen Rønnest

Analysis of Multivariate Sensor Data for Monitoring of Cultivations

PhD Thesis, July 2010

Analysis of Multivariate Sensor Data for Monitoring of Cultivations,

This report was prepared by
Nanna Petersen Rønnest

Supervisors

Krist V. Gernaey	Department of Chemical and Biochemical Engineering, DTU
Anna Eliasson Lantz	Department of Systems Biology, DTU
Stuart Stocks	Novozymes A/S, Bagsværd, Denmark

Department of Chemical and Biochemical Engineering
Center for Process Engineering and Technology (PROCESS)
Technical University of Denmark
Søltofts Plads, building 229
DK-2800 Kgs. Lyngby
Denmark

www.kt.dtu.dk
Tel: (+45) 45 25 28 00
Fax: (+45) 45 93 29 06
E-mail: kt@kt.dtu.dk

Release date:	July 2010
Category:	1 (public)
Edition:	First
Comments:	This report is part of the requirements to achieve the PhD in Chemical Engineering at the Technical University of Denmark.

Preface

The work presented in this thesis was conducted at the Center for Process Engineering and Technology (PROCESS), Department of Chemical and Biochemical Engineering, the Technical University of Denmark, from October 2006 to July 2010. The project was carried out within the framework of the Innovative Bioprocess Technology Consortium, financed by DTU, the Danish Research Council for Technology and Production Sciences, Chr. Hansen A/S, Danisco A/S and Novozymes A/S.

First of all I would like to thank my three supervisors, who each contributed with something different and valuable to the project. To Krist V. Gernaey, who provided valuable scientific input and equally important managed to create a very pleasant and productive work environment with his great support and trust. To Anna Eliasson Lantz for her very detailed evaluation of the work and serious and knowledgeable input, which has been indispensable in this project. To Stuart Stocks from Novozymes for his great interest in the project, his careful evaluation of the work and valuable input from an industrial angle. Furthermore, I would like to thank Stuart Stocks for allowing me to take part in the analysis of the size distribution data from the *Aspergillus oryzae* fermentations and for his great help during the stressful weeks of process development of the fed batch process.

I would also like to thank my former and present colleges at PROCESS and at the Center for Microbial Biotechnology (CMB), Department of Systems Biology for making the last four years so enjoyable. I am particularly grateful to Peter Ödman, who worked closely with me in this project and who tackled most of the experimental challenges in the start up of this project. Moreover, I was so lucky to work with a very talented Master student Albert E. Cervera, who contributed greatly to this project (Chapter 3 and 8). I owe many thanks to Prashant Babat, Sujata Vijay Sohoni, Stig Rattleff and Linda Olkjær Lehmann and of course to the technical staff at CMB, Jette Mortensen, Tina Johansen, Martin Nielsen and Peter Meincke for their help with the *Streptomyces coelicolor* cultivations.

Finally, I am deeply grateful to my husband Mads and daughter Iben for their great support during the last four years, particularly during the periods with nightly excursions to DTU to collect samples and in this last period of thesis-writing.

Abstract

Improved monitoring of cultivations has the potential to increase process understanding, enable immediate quality assessment, improve process control, and facilitate optimization of the process. Over the past decades various advanced sensor technologies have been developed, which can facilitate rapid estimations of important biological or chemical variables in the cultivation processes. Despite a large number of academic studies, most of the advanced sensors have not found widespread use in the industry until now. The main aim of this thesis has been to address this apparent discrepancy. The thesis consists of four case studies, which have provided critical evaluations of various sensor technologies and have served to elucidate some of the specific challenges of advanced sensor application in filamentous and industrially relevant cultivation processes.

The first two case studies were focused on quantification of filamentous morphology using off-line laser diffraction and the relation of the size distribution to the rheological properties of the cultivation broth. Firstly, the method of laser diffraction was compared to simple image analysis for quantification of clumps and pellets in *Streptomyces coelicolor* cultivation broth. The laser diffraction technique and image analysis yielded size distributions with similar shape, i.e. unimodal or bimodal distributions. The two techniques produced similar estimations of the population means, whereas the estimates of the standard deviations were generally higher using the laser diffraction technique compared to image analysis. In the second study, laser diffraction was used to quantify the morphology in industrial *Aspergillus oryzae* cultivations. Systematic changes in the size distribution were correlated to (1) the change in morphology as the cultivation progressed and (2) the difference in morphologies observed between continuously fed batches and the batches with pulse-pause feeding. Models were calibrated for prediction of the rheological parameters yield stress, consistency index and the apparent viscosity of *A. oryzae* broth based on the size distribution. Validation on an independent test set yielded a root mean square error of 1.20 Pa for the yield stress, 0.209 Pa s^n for the consistency index and 0.0288 Pa s for apparent viscosity, corresponding to R^2=0.95, R^2=0.94, and R^2=0.95 respectively.

In the third case study, in situ NIR spectroscopy was evaluated for quantification of of two important analytes: glucose and ammonium in *S. coelicolor* batch cultivations. In order to critically evaluate the ability of NIR to measure the two specific analytes, the inherent correlations in the batch process were broken by the use of semi-synthetic samples measured in situ. In addition, samples were

collected and analyzed using off-line NIR, to highlight the specific challenges of in situ monitoring. The results showed that both off-line and in situ NIR are suitable for the measurement of glucose concentration in the range 1-40 g/L. Using in situ NIR spectroscopy, a prediction error or 1.1 g/L glucose was obtained corresponding to 5.4 % of the mean value in a validation batch. For measurement of ammonium in the concentration range 40-110 mM, a large increase in prediction error was observed for in situ measurements (RMSEP= 11 mM corresponding to 16% of the mean value) compared to off-line measurements (RMSEP = 5.1 mM corresponding to 7.1 % of the mean value). This may be partly explained by a reduction in usable wavelength regions in the in situ NIR spectra compared to the spectra collected off-line, as well as increased noise in the in situ measurements.

In the fourth case study, multi-wavelength fluorescence, scanning dielectric spectroscopy (DE), on-line turbidity (OD), and simple software sensors were evaluated for the on-line estimation of biomass concentration in *S. coelicolor* fed-batch cultivations. The experiments were designed to imitate an industrial process development, thereby introducing significant variation in the process parameters. In this highly challenging system, it was shown that the more advanced sensors with multivariable output did not improve predictions of dry cell weight compared to the more simple measurements of dual frequency DE, on-line turbidity and measurements of carbon dioxide evolution rate (CER). However, the predictions of dry cell weight were greatly improved by combining selected sensors. Thus a prediction error of 1.5 g/L, corresponding to 6 % of the covered biomass range was obtained by using a combination of dual frequency DE, on-line OD and CER measurements. Furthermore, the use of multiple sensors enabled supervision of the predictions in real-time, facilitating an operator intervention such as a change between individual sensors or the collection of an off-line sample for recalibration. It was thereby demonstrated that the on-line biomass sensors fulfill the demanding requirements of process development and can be of great industrial use in both pilot plant and/or industrial production scale.

Dansk Resumé

Forbedret overvågning af kultiveringsprocesser kan potentielt forbedre forståelsen af processerne, muliggøre umiddelbar kvalitetskontrol, forbedre kontrollen af processerne og være til gavn ved processoptimering. Gennem de sidste årtier er forskellige avancerede sensorer blevet udviklet, som kan give hurtige målinger af vigtige kemiske og biologiske variable. På trods af en lang række akademiske studier, har de fleste af de avancerede sensorer ikke vundet stor udbredelse i industrien. Det vigtigste formål med denne afhandling har været at undersøge dette umiddelbare misforhold. Afhandlingen består af fire eksperimentelle studier, som har undersøgt forskellige sensor teknologier og som har afdækket nogle af de særlige udfordringer for anvendelse af avancerede sensorer i filamentøse og industrielt relevante kultiveringer.

De første to studier omhandler karakterisering af filamentøs morfologi ved hjælp af laser diffraktion, og sammenhængen mellem størrelsesfordelingen af biomasse partikler og de reologiske egenskaber. Først blev laser diffraktion sammenlignet med simpel billedanalyse af et antal prøver fra *Streptomyces coelicolor* fed batch kultiveringer. Laser diffraktion og billedanalyse gav størrelsesfordelinger med lignende overordnet form, d.v.s enten unimodal eller bimodal. Forskellen på de estimerede gennemsnitsværdier for partikelpopulationerne var i alle tilfælde under 65 μm, svarende til maximalt 22 % af den nominelle værdi. Dog var estimaterne af standardafvigelserne generelt lavere for laser diffraktion sammenlignet med billedanalyse med en forskel på mellem 18 og 94 μm. I det andet studie blev størrelsesfordelingen målt i en række prøver fra industrielle *Aspergillus oryzae* kultiveringer med laser diffraktion. Systematiske ændringer i størrelsesfordelingen kunne korreleres med (1) ændringer i morfologien med kultiveringsalderen og (2) den observerede forskel i morfologi mellem kultiveringer med kontinuert føde og 'puls-pause' føde. Modeller blev kalibreret til prædiktion af tre reologiske parametre: hvileforskydningsspændingen, viskositet og 'consistency index' på baggrund af størrelsesfordelingen. Validering på et uafhængigt testsæt resulterede i prædiktionsfejl på 1.2 Pa for hvileforskydningsspændingen, 0.209 Pa s^n for consistency index og 0.0288 Pa s for viskositeten svarende til henholdsvis R^2=0.95, R^2=0.94, og R^2=0.95.

I det tredje studie blev nær infrarød spektroskopi (NIR) brugt til at måle to vigtige analytter: glukose og ammonium in situ i *S. coelicolor* batch kultiveringer. Den naturlige korrelation mellem analytterne blev brudt vha. semi-syntestiske prøver for at sikre at de kemometriske modeller var baseret på signaler fra de to

respektive analytter. For at illustrere de særlige udfordringer ved in situ overvågning, blev alle prøver analyseret med off-line NIR spektroskopi til sammenligning. Resultaterne viste at både off-line og in situ NIR er egnet til måling af glukose koncentrationen (1-40 g/L). Modellen baseret på in situ NIR gav en prædiktionsfejl på 1.1 g/L svarende til 5.4 % af gennemsnitsværdien i et uafhængigt testsæt. Modellerne til prædiktion af ammonium koncentrationen gav langt større fejl baseret på in situ NIR spektre (RMSEP=11 mM svarende til 16% af gennemsnitsværdien) sammenlignet med modellen baseret på off-line NIR (RMSEP = 5.1 mM svarende til 7.1 % af gennemsnitsværdien). Dette kan dels forklares med øget støj i spektrene, der blev opsamlet in situ og dels ved en reduktion i det brugbare bølgelængdeområde ved in situ NIR spektroskopi.

I det fjerde studie blev en række avancerede sensorer til måling af biomasse testet i *S. coelicolor* fed batch kultiveringer. De forskellige sensorer var: 2-D fluorescens, dielektrisk spektroskopi, on-line turbiditet og simple software sensorer baseret på de traditionelt opsamlede overvågningsdata. Eksperimenterne blev designet som en industriel procesudvikling, hvorved stor variation blev introduceret i procesparametrene. Resultaterne viste at de mere avancerede sensorer ikke forbedrede biomasse målingerne sammenlignet med de mere simple sensorer såsom dielektriske spektroskopiske målinger ved to frekvenser, on-line turbiditet og måling af CO_2 i udgangsgassen i dette meget udfordrende system. Dog blev biomassemålingerne forbedret ved brug af en kombination af flere sensorer. Ved at bruge en kombination af dielektriske spektroskopiske målinger ved to frekvenser, on-line turbiditet og måling af CO_2 i udgangsgassen blev prædiktionsfejlen reduceret til 1.5 g/L svarende til 6 % af det målte biomasseområde. Desuden gav brugen af flere sensorer mulighed for overvågning af biomassemålingerne. Det blev dermed demonstreret at biomassesensorerne opfylder de høje krav i procesudviklingen. Netop i processudviklingen er der stort potentiale for anvendelse af flere og bedre sensorer i det øget overvågning her kan bidrage med en bedre karakterisering af processerne, som vil være til hjælp i procesoptimeringen og ved udvikling af kontrolstrategier.

Contents

Abbreviations:

A1-4, B1-2, C1-2	*S. coelicolor* fed batch cultivations
ANN	artificial neural networks
CER	carbon dioxide evolution rate
DCW	dry cell weight
DE	dielectic
der.	derivative
EN	electronic nose
df	degrees of freedom
FP1-2	*S. coelicolor* batch cultivations for first principles based modeling
FT	Fourier transform
GA	genetic algorithm
iPLS	interval partial least squares
LV	latent variables
LWR	locally weighted regression
MIR	mid infrared
MSC	multiplicative scatter correction
MSPC	multivariate statistical process control
MWF	multi-wavelength fluorescence
NIR	near infrared
N-PLS	multi-linear partial least squares
OD	optical density
OUR	oxygen uptake rate
PARAFAC	parallel factor analysis
PAT	process analytical technology
PCA	principal component analysis
PLS	partial least squares
RMSEC	root mean square error of calibration
RMSEP	root mean square error of prediction
RMSECV	root mean square error of cross-validation
rpm	rounds per minute
RQ	respiratory quotient
SEC	standard error of calibration
SEP	standard error of prediction
SECV	standard error of cross-validation
SNV	standard normal variate
SPC	statistical process control
SS1-5	semi-synthetic samples from *S. coelicolor* batch cultivations
ST1-2	standard *S. coelicolor* batch cultivations
SWS	spectral window selection
TOS	theory of sampling

Symbols:

A	absorbance (Chapter 3)
A	number of components (Chapter 4)
$\mathbf{B}$	regression coefficient matrix
C	capacitance (F)
C_∞	capacitance at high frequencies (F)
d	diameter
D^2	Mahalanobis distance
$\mathbf{E}, \mathbf{F}$	residual matrices
f	frequency (Hz)
f_c	characteristic frequency (Hz)
I	number of samples
J, K	number of variables
K	consistency index (Pa s^n)
k_La	liquid side oxygen mass transfer coefficient (s^{-1})
n	flow behavior index (-)
$\mathbf{P}$	X-loading matrix
$\mathbf{Q}$	Y-loading matrix
$s()$	standard deviation
S_0	singlet ground state
S_1, S_2	singlet excited electronic states
T_1	triplet excited electronic state
$\mathbf{T}$	score matrix
X	independent variable(s)
Y	dependent variable(s)
Z	test size
γ	shear rate (s^{-1})
λ	wavelength (m)
μ	viscosity (Pa s)
μ_{app}	apparent viscosity (Pa s)
$\mathbf{\Sigma}$	covariance matrix
τ	stress (Pa)
τ_y	yield stress (Pa)

Vectors are written in lower case **bold** and matrices are written in upper case **bold**. The superscript $\hat{}$ is used for estimated values, $'$ is used for transposed and a bar $\bar{}$ is used for the mean value.

Chapter 1

Introduction

Cultivation processes are widely used in industrial production for the manufacture of a range of products such as pharmaceuticals, enzymes, food ingredients etc. Cultivation processes are highly complex, involving a large number of chemical and biochemical reactions as well as heat and mass transport processes. Today, on-line monitoring and control of industrial cultivation processes mainly rely on the measurement of a limited number of chemical and physical variables such as the temperature, the pH, the exhaust gas composition and the dissolved oxygen concentration. Careful analysis of these traditional measurements can provide important information about the cultivation process. For example, the O_2 uptake and CO_2 production are tightly linked to the cellular growth and metabolism under aerobic conditions. Furthermore, the oxygen supply is often limiting in high-viscosity cultivations, and as a consequence measurements of dissolved O_2 can be used for efficient control of aerobic cultivation processes.

Access to additional on-line measurements, for example of important chemical and biochemical variables, can potentially further improve the understanding of cultivation processes, and can thereby improve to our ability to control cultivation processes. Over the past decades, new advanced technologies have been developed, which can provide more detailed information about the chemical and biological variables in the cultivation process such as substrate, product, and biomass concentrations. The recent Process Analytical Technology (PAT) initiative by the Food and Drug Administration has increased the focus on process analytical technologies by encouraging the use of advanced on-line sensors.

At present time advanced sensor technologies such as Near Infrared (NIR) spectroscopy, Dielectric (DE) spectroscopy and Multi-Wavelength Fluorescence (MWF) spectroscopy are readily available for on-line monitoring of cultivations. Numerous successful studies have been published on the utilization of these technologies for cultivation monitoring, but to our knowledge most of the technologies have not found widespread acceptance in the cultivation industry. Thus, there seems to be a discrepancy between the obvious potential of these technologies and the practical application in industry. The major aim of this work is to bridge part of the gap between the successful academic studies of the advanced sensors and industrial application. This is done by a critical evaluation of some of the most promising technologies in two challenging and industrially relevant filamentous cultivation systems. The evaluation is performed with particular focus of the specific challenges of advanced sensor application in an industrial environment.

For successful application of monitoring technology in cultivation processes, there are a number of questions to answer and challenges to be met. First of all, which sensors are suitable for measurements of for example morphology or

nutrient, metabolite or biomass concentrations? In this work various sensors have been applied: laser diffraction for quantification of filamentous morphology, NIR spectroscopy for measurement of substrate concentrations, and various biomass sensors such as MWF spectroscopy, DE spectroscopy, on-line turbidity and software sensors. Another important question to answer is: which experiments are required for the calibration of the sensors? Often cellular metabolism introduces correlations between the different analytes, such as substrate, metabolites and biomass. It is critical to address the problem of correlations in order to ensure an analyte specific sensor calibration. Furthermore, the calibration data must contain sufficient variation to obtain calibrated models which cover the desired range of process conditions, including deviations from normal operation. Two approaches are taken in this work to ensure sufficient variation in the data and elimination of correlations: (1) inclusion of semi-synthetic samples in the calibration is used in Chapter 8 for calibration of a NIR sensor and (2) introduction of variations in the process protocol used in Chapter 9 for calibration of biomass sensors. The thesis is structured into 11 chapters describing the background of the work, four experimental case studies and a discussion of the application of advanced sensors in an industrial setting. Some of the material in this thesis has been published or submitted for publication (Chapter 3, 7, 8 and 9). The papers are included as they were published, with only small modifications to ensure uniform formating.

Chapters 2-4 provide the theoretical background of the work. Chapter 2 presents a short literature review on the application of classical monitoring, MWF and dielectric spectroscopy in cultivation processes. A detailed review of NIR spectroscopy in cultivation monitoring is given in Chapter 3. Chapter 4 provides a short introduction to multivariate data analysis used in the analysis of the sensor data.

Chapter 5 provides an overview of the experimental work that has been carried out during this project, which forms the basis of the analyses described in Chapters 6 to 9. In addition, a more detailed description is given of the *Streptomyces coelicolor* cultivations.

Chapters 6 and 7 describe two case studies focused on the use of laser diffraction for quantification of filamentous morphology. Chapter 6 presents a brief comparison of the results obtained using laser diffraction and image analysis for quantification of *S. coelicolor* cell clumps and pellets. In Chapter 7, the size distribution data provided by the laser scattering technique is correlated to the rheological properties of the cultivation broth.

Chapter 8 presents a study of in situ NIR for monitoring of glucose and ammonium in *S. coelicolor* cultivations using semi-synthetic samples to break the correlations. The results are compared to off-line results in order to evaluate the particular challenges of in situ measurements.

In Chapter 9, the use of MWF, dielectric spectroscopy, on-line turbidity, and software sensors are evaluated for monitoring of the biomass concentration in *S. coelicolor* fed-batch cultivations during process development. Furthermore, the advantages of using a combination of sensors are illustrated, e.g. improved prediction and supervision of sensor performance in real-time.

Chapter 10 provides guidelines for application of on-line sensors for monitoring

of industrial cultivations based on the results described in the other chapters combined with a critical literature review.

Finally, a summary and outlook is presented in Chapter 11.

Chapter 2

Monitoring of cultivation processes

Monitoring of cultivations is important in order to control the process and ensure the desired conditions for the microorganisms. Monitoring can be performed on different levels of detail and on different time scales. Thus cultivation monitoring can range from simple observations, such as observations of a color or pH change, to a complete biochemical characterization of the medium and biomass. A distinction if often made between off-line, at-line and on-line monitoring. Off-line refers to the analysis of a sample collected manually from the bioreactor. The analysis is carried out off-site, e.g. in a laboratory, and entails a time delay ranging from hours to several days. Similarly, at-line monitoring requires manual sampling, however, the analysis is performed in close proximity to the bioreactor and can be relatively fast (in the range of minutes to hours). On-line monitoring in completely automatic and covers both in situ measurements, which are made inside the bioreactor and ex-situ measurements, where the measurements are made outside the reactor, for example in a flow through cell or by automatic collection and analysis of a sample from the bioreactor. On-line monitoring can be performed very fast with an analysis time ranging from seconds to minutes. In order to utilize the information for control of the process, the analysis should be relatively fast and preferably with-out any delay, i.e. in real-time.

This chapter presents an overview of the different on-line sensors applied for cultivation monitoring in this thesis. Firstly, classical cultivation monitoring is presented along with a few examples of software sensors based on the classical monitoring data. Furthermore, MWF and dielectric spectroscopy are presented in brief literature reviews covering the basic principles of operation and examples of application for cultivation monitoring. A separate chapter is devoted to a more in depth literature review of applications of near infrared spectroscopy in fermentation and cells cultures (Chapter 3). Finally, different methods for characterization of morphology are presented in Chapter 6.

2.1 Classical cultivation monitoring

A number of important physical and chemical variables are routinely measured on-line in both academic and industrial facilities. This includes feed rate, temperature, pH, aeration (rate and composition), electrical power consumption, stirrer speed, weight/volume, pressure, dissolved oxygen concentration, pH and exhaust gas composition (Sonnleitner, 1999).

pH and temperature are measured and automatically controlled in most industrial scale cultivation processes because both have a large effect on the growth

and metabolism of microorganisms (Sonnleitner, 1999). Most microorganisms grow optimally in a limited range of temperature and pH values. Furthermore, pH and temperature may influence metabolism, for example a change in pH may cause a switch in secondary metabolism of some organisms (Kim et al., 2007).

For aerobic processes, measurement of dissolved oxygen concentration and analysis of the exhaust gas can provide valuable information about the cultivation process. In processes of high viscosity the dissolved oxygen concentration may be one of the limiting factors. The level is often controlled, for example by adjustment of aeration rate, stirrer speed, or by adjusting the feed rate. The composition of the exhaust gas, e.g. the level of O_2 and CO_2, can be used to calculate the carbon dioxide evolution rate (CER) and the oxygen uptake rate (OUR) which are closely linked to the activity of the cells such as the growth (Sonnleitner, 1999). The respiratory quotient (RQ) can be calculated based on measurements of the O_2 and CO_2 concentration in the inlet and exhaust gas and is an important indicator of the physiological state of the cultivation process. The RQ is an example of a simple software sensor (Sonnleitner, 1999).

The measurement and control of the above mentioned variables is often a prerequisite for reproducible operation of cultivation processes. Furthermore, the widespread use of these variables for cultivation monitoring has resulted in a very large knowledge base linking these measurements to process behavior. However, the physical and chemical variables do not provide direct measurement of the biological performance of the process. Measurements of variables such as extracellular concentrations of substrates and products, biomass concentration, and intracellular levels of enzymes and RNA would provide a better characterization of the processes, which could help to increase understanding (Olsson et al., 1998; Sonnleitner, 1999). However, measurement of these chemical and biochemical variables is very challenging and is thus mainly performed off-line for example using chromatography. Firstly, the concentration levels of important chemical and biological variables are often quite low (mM range or lower). Furthermore, for on-line application, the sensor must be able to operate over long periods and in dynamic conditions (changing physical and chemical properties of the broth), be resistant to degradation or enzymatic break down and insensitive to protein absorption or surface growth (Sonnleitner, 1999). Finally, for on-line application, the sensor should not compromise the sterile environment of the bioreactor. The sensor therefore either has to be able to withstand repeated sterilization for in situ measurements or be combined with a robust and reliable sampling system (Sonnleitner, 1999).

2.2 Software sensors

Software sensors can provide estimates of desired variables and kinetic parameters based on measurements of related variables through a mathematical model (Cheruy, 1997; Sonnleitner, 1999). The input to the software sensors may be both classical monitoring data and/or measurements from more advanced sensors. Furthermore, software sensors can be based on either mechanistic models, black-box models or combinations of the two (James et al., 2001). Mechanistic models are based on mathematical descriptions of the underlying *mechanisms*, such as mass

and energy balances. Examples include a recent study by Sundström and Enfors (2008) who developed four software sensors based on simple mathematical models. The sensors estimated the biomass concentration, the specific growth rate, the oxygen transfer capacity of the bioreactor and a new variable, the $R_{O/S}$, which is the ratio between oxygen and energy substrate consumption. The calculations were based on classical monitoring variables such as the added base (biomass and specific growth rate) and measurements of oxygen concentration and inlet and exhaust gas analysis (oxygen transfer and $R_{O/S}$). A more complex sensor was developed by Jenzsch et al. (2006a) who set up a model based on simple balance equations describing the biomass concentration, substrate concentration, culture mass and the specific growth rate. The state variables were estimated based on measurements of the CER, OUR, and added base using an extended Kalman filter procedure and a simple measurement model. These estimates were then applied in a generic model control algorithm to control the specific growth rate in recombinant *E. coli* cultivations (Jenzsch et al., 2006b). One of the main advantages of the mechanistic models is the fact that they are so easily interpretable and thus have the potential to greatly improve the process knowledge for example by facilitating hypothesis testing. However, the setup of a suitable mechanistic model is dependent on some knowledge of the process and may be quite time consuming. Furthermore, depending on the complexity of the model, on-line estimations may require considerable computational power and time (Jenzsch et al., 2006b).

The black box models do not require any knowledge of the process but are based on the establishment of correlations between the measured variables and the desired variables. Neural networks and multivariate statistical tools such as principal component analysis (PCA) and partial least squares (PLS) regression are commonly used in software sensors (Jenzsch et al., 2006a; Linko et al., 1999; Zhang and Lennox, 2004). The models are more difficult to interpret compared to the mechanistic models and may only cover limited operating conditions. Examples of software sensors based on black box models include estimations of biomass concentration based on the feed rate, liquid volume and dissolved oxygen using neural networks (Chen et al., 2004) and biomass estimates based on measurements of as substrate feed rate, aeration rate, agitator power, substrate feed temperature, substrate concentration, dissolved oxygen concentration, culture volume, pH, fermenter temperature and generated heat (Zhang and Lennox, 2004). Jenzsch et al. (2006a) performed a comparative study of different model types in software sensors. They developed different software sensors for estimation of biomass concentration based on OUR, CER and base addition using neural networks, multiple linear regression, PCA and a mechanistic model. They found that the sensors using neural networks were more accurate during normal operation but that the mechanistic models were more robust during abnormal operation (Jenzsch et al., 2006a).

2.3 Fluorescence spectroscopy

Fluorescence spectroscopy is a highly sensitive and selective spectroscopic technique. As the name implies, fluorescence spectroscopy measures *fluorescence*, a phenomenon only exhibited by relatively few molecules. The detection limits are

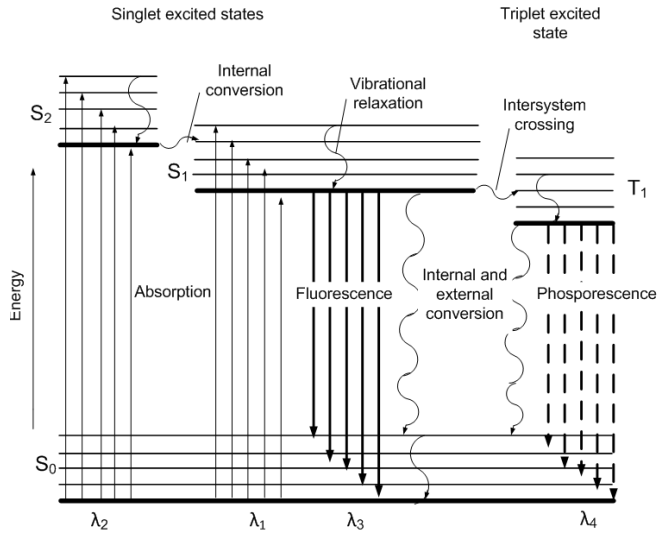


Figure 2.1: Partial energy diagram for a photoluminescent system modified from Skogg and Leary (1992).

often 1 to 3 orders of magnitude smaller than those of absorption spectroscopy; typically in the parts-per-billion range (Skogg and Leary, 1992). The existence of a number of biological fluorescent compounds has enabled the use of fluorescence spectroscopy for cultivation monitoring. This has been further promoted by the commercial availability of spectrometers for on-line measurements in bioreactors. Multi-wavelength fluorescence (MWF) spectrometers have been developed, which can measure fluorescence spectra over several excitation and emission wavelengths resulting in a three dimensional landscape (Lindemann et al., 1998; Marose et al., 1998). This facilitates the measurement of several fluorescent compounds simultaneously, which is of great use in cultivation monitoring (Marose et al., 1998).

2.3.1 Measurement principle

Photoluminescence, which covers fluorescence and phosphorescence, involves excitation of a molecule to en excited state and subsequent relaxation of the molecule accompanied by emission of light (Skogg and Leary, 1992). The phenomenon is best described through an energy diagram as depicted in Figure 2.1. A molecule in the ground state (S_0) can be excited to a higher electronic state (S_1 or S_2) by absorption of light at wavelengths centered around at λ_1 or λ_2 respectively (Figure 2.1). The electronic states are associated with different vibrational energy levels, shown as the heavy lines (ground vibrational state) and the lighter lines in Figure 2.1. S_1 and S_2 are singlet states, which means that the spin of the promoted electron is still paired with the ground state electron. However, in some cases the spin is reversed by *intersystem crossing* giving rise to the triplet state T_1. The excited molecules can return to the ground state by a combination of different steps (Skogg and Leary, 1992). A distinction can be made between the mechanisms, which involves the emission of a photon (indicated by the straight arrows) and the radiationless mechanisms (indicated by the curvy arrows in the

diagram). *Vibrational relaxation* is the loss of vibrational energy, which occurs immediately (approximately 10^{-12} s) through molecular collisions. Due to the short lifetime of the vibrationally excited molecules, fluorescence from a solution always occurs from the lowest vibrational band of an excited electronic state (Skogg and Leary, 1992). *Internal conversion* covers the intermolecular processes by which a molecule is passes from one electronic state to a lower-energy state. This is very common when the different vibrational levels from different electronic states overlap, but otherwise the process is not well understood (Skogg and Leary, 1992). *External conversion* describes the deactivation of an electronic state by energy transfer to other molecules. *Fluorescence* covers the relaxation of the molecule in an excited singlet state to one of the vibrational energy levels of the ground electronic singlet state. This transmission is associated with the emission of light. The wavelength of the emitted light is centered around λ_3 depending on the vibrational energy in the ground electronic state. Due to the efficiency of the vibrational relaxation, the emission band is shifted towards the lower frequencies. This is referred to as the Stokes shift (Skogg and Leary, 1992). Finally, *phosphorescence* describes the relaxation of a molecule in a excited triplet state accompanied by the emission of a photon. The transition from a triplet to a singlet state is much less probable than a singlet-singlet transition and thus the average lifetime of an excited triplet state is in the range 10^{-4} to 10 s or more.

The *quantum yield* or *quantum efficiency* is defined as the ratio between the number of molecules that luminesce to the total number of excited molecules (Skogg and Leary, 1992). *Quenching* refers to all of the various processes that result in a reduction of the true fluorescence efficiency being reduced to below unity, for example the processes mentioned above (Parker and Rees, 1962).

The ability of a compound to fluoresce is dependent on both the molecular structure and the chemical environment. It has been found that compounds containing aromatic functional groups with low-energy $\pi \rightarrow \pi*$ transition levels often give rise to intense fluorescence (Skogg and Leary, 1992). Furthermore, compounds which contain aliphatic and alicyclic carbonyl structures or highly conjugated double bond structures may also fluoresce, but this is not as common as for the aromatic compounds (Skogg and Leary, 1992). Important fluorescent compounds for monitoring of cultivation processes include NAD(P)H, ATP, aromatic amino acids (e.g. tryptophan) and several vitamins (e.g. riboflavin and pyridoxine). Fluorescence is also dependent on a number of external factors. Decreased temperature and increased viscosity tend to reduce the number of collisions and thus improve fluorescence. pH and polarity of the solvent can affect the fluorescence spectra through interactions with the fluorescent compounds. Finally, dissolved oxygen quenches fluorescence, commonly by promoting intersystem crossing (Valeur, 2001).

In addition, the measurements of fluorescence may be subject to *inner filter effects*, which cover the reduction of the intensity of the observed light due to absorption or scattering of the exciting or emitted light (Li and Humphrey, 1992; Skogg and Leary, 1992). This may be caused by the presence of other absorbing or scattering species in the sample or by *self-absorption*. Self-absorption can occur when the emission band overlaps with an absorption band (Skogg and Leary, 1992).

The above mentioned factors may all influence fluorescence spectroscopy for bioprocess monitoring. Li and Humphrey (1992) studied the effect of probe position, agitation and aeration rates, pH, temperature, dissolved oxygen, inner filter effects on the fluorescence signals from a bioreactor system. They found that the inner filter effects, temperature, probe position, aeration and agitation rates were particularly important in the quantification of the fluorescent signals from the cultivation. In addition, they found that the pH effect was mostly important in quantifying tryptophan and pyridoxine fluorescence. In their study, the dissolved oxygen concentration did not pose a serious problem for the the quantification of either tryptophan, pyridoxine, NADH and riboflavin dissolved in buffer. In addition, Li and Humphrey (1992) developed models which describe the effect of pH, temperature and inner filter effects on the back scattered fluorescence signals and which can be used to correct the fluorescent signals.

2.3.2 Measurement of biomass concentration

A number of studies have used MWF for estimation of biomass concentration in a various cultivation processes such as cell cultures (Hisiger and Jolicoeur, 2005; Teixeira et al., 2009), bacteria (Clementschitsch et al., 2005; Hagedorn et al., 2003; Lantz et al., 2006; Skibsted et al., 2001), yeast (Haack et al., 2007; Hisiger, 2005; Ödman et al., 2009; Surribas et al., 2006a), and filamentous organisms (Haack et al., 2007; Ödman, 2009). Different fluorophores have been identified such as NADH, flavins, tryptophan, pyridoxine in the cultivation broths. In the majority of the cited studies, it was not possible to correlate the biomass concentration to one single fluorophore. Instead chemometric methods such as PLS, neural networks, N-PLS and parallel factor analysis (PARAFAC) in combination with a regression method were employed for estimation of the biomass concentration.

In some processes with varying physiological states, it has not been possible to calibrate one single model for the prediction of biomass over the entire process (Ödman et al., 2009; Surribas et al., 2006b). For example, excellent models have been obtained for prediction of biomass in yeast cultivations during exponential growth on glucose, but the same models failed in the ethanol consumption phase (Ödman et al., 2009). In those cases, segmented models can be used to provide predictions of biomass concentration over the whole process (Ödman et al., 2009). In case of variation between batches, for example due to variation of medium composition, it has similarly been difficult to estimate one global model for prediction under all conditions (Lantz et al., 2006; Ödman et al., 2009). However, in other cases it has been possible to calibrate models which cover substantial process variation (Ödman, 2009)

2.3.3 Measurement of other variables

Multi-wavelength fluorescence has been used in a number of applications for the measurement of other fluorescent compounds such as different protein products (Haack et al., 2007; Hisiger, 2005; Mortensen and Bro, 2006; Surribas et al., 2006b), alkaloids (Boehl et al., 2003; Hisiger and Jolicoeur, 2005) and polymyxin, an antibiotic (Lantz et al., 2006). In most of the above mentioned studies, a specific signal from the fluorescent analyte was identified, predictions of the ana-

lytes were often improved by chemometric modeling of a larger wavelength area (Boehl et al., 2003; Haack et al., 2007; Skibsted et al., 2001). In addition, MWF spectroscopy has also been used for estimation of concentration levels of other non-fluorescent but correlated analytes. This includes succinate, nitrate (Skibsted et al., 2001), glycerol, methanol (Surribas et al., 2006b), glucose and ethanol (Ödman et al., 2009).

By enabling measurement of important biological analytes, such as NAD(P)H, in very small concentrations, MWF has the potential to measure more complex process variables such as the physiological or metabolic state of the cells. MWF has been used to measure the cell activity (Hisiger and Jolicoeur, 2005) and respiometric activity (Skibsted et al., 2001), both by measurement of NADH. Moreover, measurements of fluorescence can indicate phase shifts during cultivations such the onset of the feeding phase (Haack et al., 2007) and substrate exhaustion (Hagedorn et al., 2003). Hantelmann et al. (2006) monitored changes in metabolic state connected to ethanol production in *S. cerevisiae* and used this to control the feed rate to avoid ethanol production. The system efficiently eliminated the ethanol production and thus increased the biomass yield accordingly (Hantelmann et al., 2006).

2.4 Dielectric spectroscopy

Dielectric spectroscopy has been used for measurements of biomass concentration since Harris et al. (1987) showed that the radiofrequency dielectric properties of microbial suspensions are a direct and monotonic function of the radius and volume fraction of the particles constituting the suspended phase. A number of studies have been published on the topic including a number of applications with unicellular microbial systems such as yeast (Austin et al., 1994; Dabros et al., 2009b; November and Van Impe, 2000; Xiong et al., 2008) and lactic acid bacteria (Arnoux et al., 2005), animal and plant cell culture systems and a few applications with filamentous organisms (Fehrenbach et al., 1992; Kiviharju et al., 2007; Neves et al., 2000; Sarra et al., 1996). By far, the biggest industrial application of dielectric spectroscopy for biomass monitoring is the brewing industry (Kiviharju et al., 2008). Dielectric spectroscopy are by some viewed upon as the basic workhorse of on-line biomass measurements (Dabros et al., 2009a).

2.4.1 Measurement principle

Dielectric spectroscopic methods are based on the measurement of a material's response to an applied electric field (Dabros et al., 2009b). This typically includes measurements of the capacitance and conductance as a function of the applied field frequency (Dabros et al., 2009b; Markx and Davey, 1999). The capacitance quantifies the ability to store charge (measured in F), while the conductance quantifies the ability to conduct electric charge (measured in S). The capacitance and conductance of a cell suspension depends on the geometry of the electrodes used to measure them (Davey et al., 1993). To model physical process the values can be converted to exclude the effect from the probe geometry (Davey et al., 1993). The permittivity (in F/m) is the capacitance normalized to take into

account the probe geometry, and correspondingly the conductivity (in S/m) is the normalized conductance. The permittivity is often given as the permittivity relative to the permittivity of vacuum.

In living cells, the cytoplasma is surrounded by the plasma membrane, which has a lipid bilayer structure. The cytoplasma is conducting while the plasma membrane is not. When an electric field is applied, the ions will migrate towards the electrodes. However, since the plasma membrane is not conducting, this will result in an accumulation of charge at the side of the cells. The cells will thus become polarized and act as small capacitors (Dabros et al., 2009b). The polarization is dependent on the frequency of the applied field. At low frequencies there will be sufficient time for charge build-up in the cells, but above a certain frequency, the change in the electric field will be too fast for the polarization to occur. This step change in the polarization ability of the cells is called the beta-dispersion (Dabros et al., 2009b) (Figure 2.2). The beta-dispersion is commonly described by the Cole-Cole equation (Cole and Cole, 1941), given by

$$C(f) = \frac{\Delta C \left(1 + \left(\frac{f}{f_c}\right)^{(1-\alpha)} \sin\left(\frac{\pi}{2}\alpha\right)\right)}{\left(1 + \left(\frac{f}{f_c}\right)^{2-2\alpha} + 2\left(\frac{f}{f_c}\right)^{(1-\alpha)} \sin\left(\frac{\pi}{2}\alpha\right)\right)} + C_\infty \tag{2.1}$$

where f is the frequency, ΔC is the difference in capacitance, C_∞ is the capacitance at high frequencies, f_c is the characteristic frequency and α is an empirical fitting parameter (Figure 2.2). For constant cell radius and membrane properties ΔC is proportional to the volume fraction of biological cells in the suspension with intact plasma membranes. The use of dielectric spectroscopy for quantification of biomass in cultivations is based on this relationship. Ideally, the characteristic frequency is related to the morphology and average size of the cells (Cannizzaro et al., 2003; Siano, 1997; Yardley et al., 2000). The rationale for this is that smaller cells polarize faster than large cells, thus resulting in a larger value of the characteristic frequency for small cells (Dabros et al., 2009b). The fitting parameter α, which has the effect of broadening the dispersion, has been reported to be related to the cell shapes and sizes (Markx et al., 1991b; Siano, 1997), the morphology of the extra-cellular spaces (Ivorra et al., 2005), mobility of the membranous proteins (Harris and Kell, 1985; Kell and Harris, 1985), the surface topology (Siano, 1997), and the fractal nature of the dielectric relaxation (Dabros et al., 2009b; Ryabov and Feldman, 2002). For further details of the theoretical modeling of the capacitance spectra readers are referred to Dabros et al. (2009b), who provides a description of the Cole-Cole equation and the parameter estimation based on the dielectric spectra.

2.4.2 Measurement of biomass concentration

The measurements of capacitance is influenced by a large number of different parameters of the cells and the surrounding medium and therefore requires calibration in order to provide estimates of the biomass concentration.

Most of the published studies estimate the biomass concentration based on linear regression on the measurements of the capacitance at a single frequency

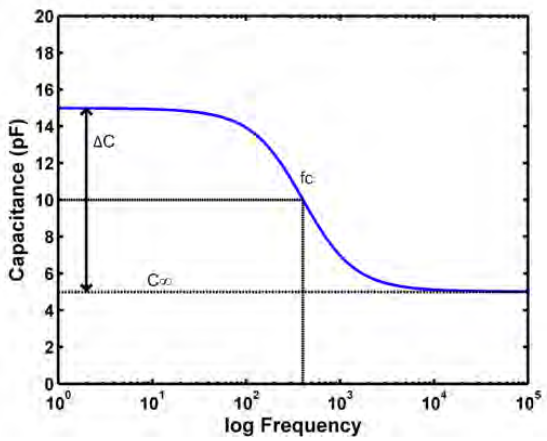


Figure 2.2: The beta dispersion. Modified from Dabros et al. (2009b).

(Ducommun et al., 2002; Guan et al., 1998; Harris et al., 1987) or the difference in capacitance between two frequencies (Dabros et al., 2009b; Fehrenbach et al., 1992; Sarra et al., 1996). By using the difference between the measurements at two frequencies it is possible to correct for instrumental drifts and fluctuations since these artifacts have the same influence on the signals at high and low frequencies (Markx et al., 1991a).

Recently, scanning over a frequency range has been applied in bioprocesses (Ansorge et al., 2007; Cannizzaro et al., 2003; Dabros et al., 2009b). In addition to the change in ΔC, the shape of the resulting spectra provides additional information concerning for example the cell size and its distribution, cell morphology, cell physiology and the conductance of the medium. The spectral data can be analyzed by fitting the theoretical Cole-Cole equation (Equation 2.1) or by the use of empirical multivariate statistical tools such as PLS and artificial neural networks (Dabros et al., 2009b; Nicholson et al., 1996). Dabros et al. (2009b) have made a comparison between linear regression based on dual frequency measurements, PLS regression based on the capacitance spectrum between 370 kHz and 15.56 MHz, and fitting to the theoretical Cole-Cole model. For prediction of biomass concentration, it was found that both linear regression and PLS regression were more robust and outperformed the Cole-Cole model in the estimation of the biomass concentration. Under noisy conditions in small bioreactors, PLS was more robust and gave lower prediction errors compared to the linear regression. In addition, the mean cell diameter was estimated using Cole-Cole model and PLS. The most accurate results were obtained with PLS.

A number of studies have explored the use of dielectric spectroscopy for the monitoring of biomass concentration in unicellular fermentation systems such as animal cell cultures, yeast and bacteria. For the bacterial and yeast fermentations it is generally agreed that dielectric spectroscopy is suitable for on-line monitoring of viable or total biomass (Arnoux et al., 2005; Austin et al., 1994; Dabros et al., 2009b; Fehrenbach et al., 1992; Harris et al., 1987; Markx et al., 1991a; Morita et al., 1999; November and Van Impe, 2000; Xiong et al., 2008). Kiviharju et al. (2007) tested dielectric spectroscopy for monitoring of biomass (OD

and dry cell weight) on a number of different bacterial and yeast fermentations. They found correlation coefficients around or above 0.9, which covered some discrepancies between the off-line and on-line measurements. In comparison to a commercial OD NIR probe, they concluded, contrarily to most other published papers, that the commercial OD probe is a more reliable method for on-line estimation of biomass concentration whereas the on-line capacitance probe requires further improvements. In animal cell cultures good agreement has generally been found between capacitance measurements and different off-line measures of the viable and the total amount of biomass during the growth phase, whereas some discrepancy has been observed when the viability of the cells is decreased (Ansorge et al., 2007; Cannizzaro et al., 2003; Cerckel et al., 1993; Ducommun et al., 2002; Guan et al., 1998). As an example, Cannizzaro et al. (2003) observed that the percentage error of prediction of viable cells in the validation batch varied between 9 % in the growth phase, to 30 % when including the last part of the cultivation where nutrients were limiting and the number of viable cells decreased. Similarly, Noll and Biselli (1998) found some discreapancy in the corralation between the capacitance signal and the viable cell concentration, which was attributed to differences in cell size and the capacitance per unit membrane area. Instead the capacitance signal correlated very well to the glutamine consumption rate, which is a measure of the metabolic activity (Noll and Biselli, 1998).

Few studies have been published concerning monitoring of the biomass concentration of filamentous organisms with dielectric spectroscopy. In the early nineties, it was demonstrated that capacitance measurements are suitable for precise biomass measurements of filamentous as well as unicellular cultures (Ciureanu et al., 1997; Fehrenbach et al., 1992; Mishima et al., 1991). In a study of *Saccharopolyspora erythraea*, Sarra et al. (1996) found a good correspondence between the biomass concentration and the capacitance measurements in the growth phase despite different morphologies of the filamentous organisms. A different correlation between the biomass concentration and the capacitance measurements was found in the biomass decline phase. It was suggested that this difference is influenced by the amount of cell debris present in the broth, which was supported by rheological measurements (Sarra et al., 1996). Krairak et al. (2000) used dielectric spectroscopy as a measure of biomass concentration in order to estimate the specific growth rate in *Monascus* cultivations. Neves et al. (2000) monitored biomass concentration using dielectric spectroscopy in industrial *Streptomyces clavuligerus* fermentations and obtained an R^2 value of 0.998 between dry cell weight up to 20 g/L in an off-line calibration. On-line, a good agreement between the dry cell weight measurements was observed up to 80 h covering the exponential, transition and stationary growth phase, in which only small amounts of dead biomass were present. Thus in the first part of the fermentation, a linear regression between the DCW and the capacitance measurements resulted in an R^2 value of 0.98 (Neves et al., 2000). The same group later tested at new probe type (annular capacitance probe) in similar fermentations and obtained an improved signal to noise ratio (Ferreira et al., 2005). In the previously mentioned study be Kiviharju et al. (2008), dielectric spectroscopy was tested for biomass monitoring in *Streptomyces peucetius* cultivations. The study showed poor correlations between the off-line biomass measurements and the capacitance measurements.

In conclusion, dielectric spectroscopy has been shown to be a promising method for estimation of live biomass in both unicellular and filamentous fermentations. However, the relatively low number of studies and the mixed conclusions concerning measurements of filamentous organisms indicate some particular challenges in these systems.

2.4.3 Indicator of physiological state, cell size and morphology

In addition to biomass concentration, the dielectric spectroscopic measurements are related to the cell physiology, cell size and the cell morphology. Thus dielectric spectroscopy have the potential to provide on-line information concerning these important parameters in addition of the biomass concentration. Monitoring of capacitance at dual frequencies have for example been used to monitor sporulation of *Bacillus thuringiens* (Sarrafzadeh et al., 2005) and the cell cycle of synchronized suspensions of a budding yeast mutant over several cell cycles (Gheorghiu and Asami, 1998). With scanning dielectric spectroscopy, additional information is available in the other parameters of the Cole-Cole equation which can be fitted to the spectra, namely the characteristic frequency, f_c, and the the relaxation constant, α. The characteristic frequency is mainly affected by the cell size, its distribution and conductance of the medium. A few studies have been able to relate the characteristic frequency to the cell size (Ansorge et al., 2007; Cannizzaro et al., 2003; Dabros et al., 2009b). Maskow et al. (2008) found that statistical analysis of the parameters of the Cole-Cole equation could accurately describe the formation of cellular lipid inclusions in *Arxula adeninivorans*. Similarly, measurements of the characteristic frequency and the capacitance of the medium can be used to estimate the conductivity of the cytoplasm, which has been shown to be correlated to the oil content in the mycelia of *Mortierella alpina* (Higashiyama et al., 1999).

The usability of these parameters for the monitoring of cell size and morphology, and in particular a quantitative determination of the two, depends on the accuracy of the parameter estimates. Firstly, noise in the spectra will influence the accuracy of the estimates and hamper the detection of the often small changes cell size and morphology. Secondly, it may be difficult to separate the different effects of changes in the conductance of the medium, membrane properties of the cells, the cell size and distribution and the cell morphology. Finally, overlapping relaxations resulting from non-spherical morphologies may hinder the deconvolution of the spectra (Siano, 1997), thus limiting the application of the dielectric spectroscopy for monitoring of cell size and morphology.

2.4.4 Advanced applications of dielectric spectroscopy

Biomass predictions by dielectric spectroscopy has been used in a number of advanced applications e.g. for feedback control. Markx et al. (1991a) introduced the 'permittistat', in which the the steady state biomass level in a continuous yeast fermentation was controlled by dielectric spectroscopy. The permittistat has later been used in a number of studies - particularly yeast fermentations (Austin et al., 1994; Davey et al., 1996; Markx and Kell, 1995). More recently, Mas et al. (2001) used biomass concentration estimated using dielectric spectroscopy to control the

flow rate in a flocculating yeast fermentation, where biomass measurements are normally hampered by the difficultly of obtaining a representative sample. Similarly, the specific growth rate has been controlled based on estimations obtained using dielectric spectroscopy (Krairak et al., 2000; November and Van Impe, 2000, 2002). Noll and Biselli (1998) used dielectric spectroscopy to estimate the glutamine consumption rate and used this in an automatic control look to control the glutamine at a constant level.

Dielectric spectroscopy has also been used in combination with other sensors to provide additional information concerning the system. By combination of microcalorimetry and dielectric spectroscopy in an animal cell culture, Guan et al. (1998) provided an on-line estimate of the specific heat flow rate, which is an indicator of total metabolic rate and thus can be useful in the formulation of a control procedure for animal cell culture systems. Clementschitsch et al. (2005) were able to improve the estimation of important process variables (such as total amount of biomass, total cell number, percentage of dead cells, recombinant product, product formation rate, and plasmid copy number) in recombinant *E. coli* cultures by combination of dielectric spectroscopy with MWF spectroscopy. Dabros et al. (2009a) combined dielectric spectroscopy with mid-infrared spectroscopy for prediction of biomass and metabolite concentrations. The concentration estimates where reconciled with off-gas analysis and base addition measurements based on the closure of elemental and mass balances (Dabros et al., 2009b). Using this procedure, it was possible to reduce the prediction errors caused by instrument and process drifts and thus provide reliable information regarding biomass and metabolite concentrations in real-time.

Bibliography

Ansorge S, Esteban G, Schmid G. 2007. On-line monitoring of infected Sf-9 insect cell cultures by scanning permittivity measurements and comparison with off-line biovolume measurements. Cytotechnology 55:115–24.

Arnoux AS, Preziosi-Belloy L, Esteban G, Teissier P, Ghommidh C. 2005. Lactic acid bacteria biomass monitoring in highly conductive media by permittivity measurements. Biotechnology Letters 27:1551–1557.

Austin GD, Watson RWJ, D'Amore T. 1994. Studies of online viable yeast biomass with a capacitance biomass monitor. Biotechnology and Bioengineering 43:337–41.

Boehl D, Solle D, Hitzmann B, Scheper T. 2003. Chemometric modelling with two-dimensional fluorescence data for *Claviceps purpurea* bioprocess characterization. Journal of Biotechnology 105:179–188.

Cannizzaro C, Guegerli R, Marison I, von Stockar U. 2003. On-line biomass monitoring of CHO perfusion culture with scanning dielectric spectroscopy. Biotechnology and Bioengineering 84:597–610.

Cerckel I, Garcia A, Degouys V, Dubois D, Fabry L, Miller O. 1993. Dielectric spectroscopy of mammalian cells. 1. Evaluation of the biomass of HeLa- and

CHO cells in suspension by low-frequency dielectric spectroscopy. Cytotechnology 13:185–93.

Chen LZ, Nguang SK, Li XM, Chen XD. 2004. Soft sensors for on-line biomass measurements. Bioprocess and Biosystems Engineering 26:191–195.

Cheruy A. 1997. Software sensors in bioprocess engineering. Journal of Biotechnology 52:193–199.

Ciureanu M, Levadoux W, Goldstein S. 1997. Electrical impedance studies on a culture of a newly discovered strain of *Streptomyces*. Enzyme and Microbial Technology 21:441–449.

Clementschitsch F, Kern J, Pötschacher F, Bayer K. 2005. Sensor combination and chemometric modelling for improved process monitoring in recombinant *E. coli* fed-batch cultivations. Journal of Biotechnology 120:183–197.

Cole KS, Cole RH. 1941. Dispersion and absorption in dielectrics I. Alternating current characteristics. Journal of Chemical Physics 9:341–351.

Dabros M, Amrhein M, Bonvin D, Marison IW, von Stockar U. 2009a. Data reconciliation of concentration estimates from mid-infrared and dielectric spectral measurements for improved on-line monitoring of bioprocesses. Biotechnology Progress 25:578–588.

Dabros M, Dennewald D, Currie DJ, Lee MH, Todd RW, Marison IW, von Stockar U. 2009b. Cole-Cole, linear and multivariate modeling of capacitance data for on-line monitoring of biomass. Bioprocess and Biosystems Engineering 32:161–173.

Davey CL, Markx GH, Kell DB. 1993. On the dielectric method of monitoring cellular viability. Pure and Applied Chemistry 65:1921–6.

Davey H, Davey C, Woodward A, Edmonds A, Lee A, Kell D. 1996. Oscillatory, stochastic and chaotic growth rate fluctuations in permittistatically controlled yeast cultures. Bio Systems 39:43–61.

Ducommun P, Kadouri A, Von Stockar U, Marison IW. 2002. On-line determination of animal cell concentration in two industrial high-density culture processes by dielectric spectroscopy. Biotechnology and Bioengineering 77:316–323.

Fehrenbach R, Comberbach M, Petre JO. 1992. On-line biomass monitoring by capacitance measurement. Journal of Biotechnology 23:303–14.

Ferreira AP, Vieira LM, Cardoso JP, Menezes JC. 2005. Evaluation of a new annular capacitance probe for biomass monitoring in industrial pilot-scale fermentations. Journal of Biotechnology 116:403–409.

Gheorghiu E, Asami K. 1998. Monitoring cell cycle by impedance spectroscopy: experimental and theoretical aspects. Bioelectrochemistry and Bioenergetics 45:139–143.

Guan Y, Evans PM, Kemp RB. 1998. Specific heat flow rate: an online monitor and potential control variable of specific metabolic rate in animal cell culture that combines microcalorimetry with dielectric spectroscopy. Biotechnology and Bioengineering 58:464–477.

Haack MB, Eliasson Lantz A, Mortensen PP, Olsson L. 2007. Chemometric analysis of in-line multi-wavelength fluorescence measurements obtained during cultivations with a lipase producing *Aspergillus oryzae* strain. Biotechnology and Bioengineering 96:904–913.

Hagedorn A, Legge RL, Budman H. 2003. Evaluation of spectrofluorometry as a tool for estimation in fed-batch fermentations. Biotechnology and Bioengineering 83:104–111.

Hantelmann K, Kollecker M, Hüll D, Hitzmann B, Scheper T. 2006. Two-dimensional fluorescence spectroscopy: A novel approach for controlling fed-batch cultivations. Journal of Biotechnology 121:410–417.

Harris CM, Kell DB. 1985. On the dielectrically observable consequences of the diffusional motions of lipids and proteins in membranes. 2. Experiments with microbial cells, protoplasts and membrane vesicles. European Biophysics Journal 13:11–24.

Harris CM, Todd RW, Bungard SJ, Lovitt RW, Morris JG, Kell DB. 1987. Dielectric permittivity of microbial suspensions at radio frequencies: a novel method for the real-time estimation of microbial biomass. Enzyme and Microbial Technology 9:181–186.

Higashiyama K, Sugimoto T, Yonezawa T, Fujikawa S, Asami K. 1999. Dielectric analysis for estimation of oil content in the mycelia of *Mortierella alpina*. Biotechnology and Bioengineering 65:537–541.

Hisiger M S; Jolicoeur. 2005. A multiwavelength fluorescence probe: is one probe enough for on-line monitoring of recombinant protein production and biomass activity. Journal of Biotechnology 117:325–336.

Hisiger S, Jolicoeur M. 2005. Plant cell culture monitoring using an in situ multiwavelength fluorescence probe. Biotechnology Progress 21:580–589.

Ivorra A, Genesca M, Sola A, Palacios L, Villa R, Hotter G, Aguilo J. 2005. Bioimpedance dispersion width as a parameter to monitor living tissues. Physiological Measurement 26:S165–S173.

James S, Legge R, Budman H. 2001. Comparative study of black-box and hybrid estimation methods in fed-batch fermentation. Journal of Process Control 12:113–121.

Jenzsch M, Simutis R, Eisbrenner G, Stueckrath I, Lübbert A. 2006a. Estimation of biomass concentrations in fermentation processes for recombinant protein production. Bioprocess and Biosystems Engineering 29:19–27.

Jenzsch M, Simutis R, Lübbert A. 2006b. Generic model control of the specific growth rate in recombinant *Escherichia coli* cultivations. Journal of Biotechnology 122:483–493.

Kell D, Harris C. 1985. On the dielectrically observable consequences of the diffusional motions of lipids and proteins in membranes. 1. Theory and overview. European biophysics journal 12:181–97.

Kim YJ, Song JY, Moon MH, Smith CP, Hong SK, Chang YK. 2007. pH shock induces overexpression of regulatory and biosynthetic genes for actinorhodin production in *Streptomyes coelicolor* A3(2). Applied Microbial Biotechnology 76:1119–1130.

Kiviharju K, Salonen K, Moilanen U, Eerikaeinen T. 2008. Biomass measurement online: the performance of in situ measurements and software sensors. Journal of Industrial Microbiology and Biotechnology 35:657–665.

Kiviharju K, Salonen K, Moilanen U, Meskanen E, Leisola M, Eerikaeinen T. 2007. On-line biomass measurements in bioreactor cultivations: comparison study of two on-line probes. Journal of Industrial Microbiology and Biotechnology 34:561–566.

Krairak S, Yamamura K, Irie R, Nakajima M, Shimizu H, Chim-Anage P, Yongsmith B, Shioya S. 2000. Maximizing yellow pigment production in fed-batch culture of *Monascus* sp. Journal of Bioscience and Bioengineering 90:363–367.

Lantz AE, Jørgensen P, Poulsen E, Lindemann C, Olsson L. 2006. Determination of cell mass and polymyxin using multi-wavelength fluorescence. Journal of Biotechnology 121:544–554.

Li JK, Humphrey AE. 1992. Factors affecting culture fluorescence when monitoring bioreactors. Journal of Fermentation and Bioengineering 74:104–111.

Lindemann C, Marose S, Nielsen HO, Scheper T. 1998. 2-dimensional fluorescence spectroscopy for on-line bioprocess monitoring. Sensors and Actuators B: Chemical 51:273–277.

Linko S, Zhu YH, Linko P. 1999. Applying neural networks as software sensors for enzyme engineering. Trends in Biotechnology 17:155–162.

Markx GH, Davey CL. 1999. The dielectric properties of biological cells at radiofrequencies: applications in biotechnology. Enzyme and Microbial Technology 25:161–171.

Markx GH, Davey CL, Kell DB. 1991a. The permittistat a novel type of turbidostat. Journal of General Microbiology 137:737–744.

Markx GH, Davey CL, Kell DB. 1991b. To what extent is the magnitude of the Cole-Cole α of the β-dielectric dispersion of cell-suspensions explicable in terms of the cell-size distribution. Bioelectrochemistry and Bioenergetics 25:195–211.

Markx GH, Kell DB. 1995. Use of dielectric permittivity for the control of the biomass level during biotransformations of toxic substrates in continuous culture. Biotechnology Progress 11:64–70.

Marose S, Lindemann C, Scheper T. 1998. Two-dimensional fluorescence spectroscopy: a new tool for on-line bioprocess monitoring. Biotechnology Progress 14:63–74.

Mas S, Ossard F, Ghommidh C. 2001. On-line determination of flocculating *Saccharomyces cerevisiae* concentration and growth rate using a capacitance probe. Biotechnology Letters 23:1125–1129.

Maskow T, Rollich A, Fetzer I, Yao J, Harms H. 2008. Observation of non-linear biomass-capacitance correlations: reasons and implications for bioprocess control. Biosensors and Bioelectronics 24:123–8.

Mishima K, Mimura A, Takahara Y. 1991. On-line monitoring of cell concentrations during yeast cultivation by dielectric measurements. Journal of Fermentation and Bioengineering 72:296–9.

Morita S, Umakoshi H, Kuboi R. 1999. Characterization and on-line monitoring of cell disruption and lysis using dielectric measurement. Journal of Bioscience and Bioengineering 88:78–84.

Mortensen PP, Bro R. 2006. Real-time monitoring and chemical profiling of a cultivation process. Chemometrics and Intelligent Laboratory Systems 84:106–113.

Neves AA, Pereira DA, Vieira LM, Menezes JC. 2000. Real time monitoring biomass concentration in *Streptomyces clavuligerus* cultivations with industrial media using a capacitance probe. Journal of Biotechnology 84:45–52.

Nicholson DJ, Kell DB, Davey CL. 1996. Deconvolution of the dielectric spectra of microbial cell suspensions using multivariate calibration and artificial neural networks. Bioelectrochemistry and Bioenergetics 39:185–93.

Noll T, Biselli M. 1998. Dielectric spectroscopy in the cultivation of suspended and immobilized hybridoma cells. Journal of Biotechnology 63:187–198.

November EJ, Van Impe JF. 2000. Evaluation of on-line viable biomass measurements during fermentations of *Candida utilis*. Bioprocess Engineering 23:473–477.

November EJ, Van Impe JF. 2002. The tuning of a model-based estimator for the specific growth rate of *Candida utilis*. Bioprocess and Biosystems Engineering 25:1–12.

Ödman P, 2009. On-line monitoring of Microbial Cultivation Processes using Near Infrared Spectroscopy and Multi-Wavelength Fluorescence. Ph.D. thesis, Technical University of Denmark, Kgs. Lyngby, Denmark.

Ödman P, Johansen CL, Olsson L, Gernaey KV, Lantz AE. 2009. On-line estimation of biomass, glucose and ethanol in *Saccharomyces cerevisiae* cultivations using in-situ multi-wavelength fluorescence and software sensors. Journal of Biotechnology 144:102–112.

Olsson L, Schulze U, Nielsen J. 1998. On-line bioprocess monitoring - an academic discipline or an industrial tool? TrAC Trends in Analytical Chemistry 17:88–95.

Parker CA, Rees WT. 1962. Fluorescence spectrometry. a review. The Analyst 87:83–111.

Ryabov YE, Feldman Y. 2002. Novel approach to the analysis of the non-Debye dielectric spectrum broadening. Physica A: Statistical Mechanics and its Applications 314:370–378.

Sarra M, Ison AP, Lilly MD. 1996. The relationships between biomass concentration, determined by a capacitance-based probe, rheology and morphology of *Saccharopolyspora erythraea* cultures. Journal of Biotechnology 51:157–165.

Sarrafzadeh MH, Belloy L, Esteban G, Navarro JM, Ghommidh C. 2005. Dielectric monitoring of growth and sporulation of *Bacillus thuringiensis*. Biotechnology Letters 27:511–517.

Siano SA. 1997. Biomass measurement by inductive permittivity. Biotechnology and Bioengineering 55:289–304.

Skibsted E, Lindemann C, Roca C, Olsson L. 2001. On-line bioprocess monitoring with a multi-wavelength fluorescence sensor using multivariate calibration. Journal of Biotechnology 88:47–57.

Skogg DA, Leary JJ. 1992. Principles of Instrumental Analysis. Fort Worth, USA: Saunders College Publishing, 4th edition.

Sonnleitner B. 1999. Instrumentation of biotechnological processes. Advances in Biochemical Engineering Biotechnology 66:1–64.

Sundström H, Enfors SO. 2008. Software sensors for fermentation processes. Bioprocess and Biosystems Engineering 31:145–152.

Surribas A, Amigo JM, Coello J, Montesinos JL, Valero F, Maspoch S. 2006a. Parallel factor analysis combined with PLS regression applied to the on-line monitoring of *Pichia pastoris* cultures. Analytical and Bioanalytical Chemistry 385:1281–1288.

Surribas A, Geissler D, Gierse A. 2006b. State variables monitoring by in situ multi-wavelength fluorescence spectroscopy in heterologous protein production by *Pichia pastoris*. Journal of Biotechnology 124:412–420.

Teixeira A, Portugal C, Carinhas N, Dias J, Crespo J, Alves P, Carrondo M, Oliveira R. 2009. In situ 2D fluorometry and chemometric monitoring of mammalian cell cultures. Biotechnology and Bioengineering 102:1098–1106.

Valeur B. 2001. Molecular fluorescence. Weinheim: Wiley-VCH Verlag.

Xiong ZQ, Guo MJ, Guo YX, Chu J, Zhuang YP, Zhang SL. 2008. Real-time viable-cell mass monitoring in high-cell-density fed-batch glutathione fermentation by *Saccharomyces cerevisiae* T65 in industrial complex medium. Journal of Bioscience and Bioengineering 105:409–413.

Yardley JE, Todd R, Nicholson DJ, Barrett J, Kell DB, Davey CL. 2000. Correction of the influence of baseline artefacts and electrode polarisation on dielectric spectra. Bioelectrochemistry 51:53–65.

Zhang H, Lennox B. 2004. Integrated condition monitoring and control of fed-batch fermentation processes. Journal of Process Control 14:41–50.

Chapter 3

Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation

The contents of this chapter has been published in Cervera AE, Petersen N, Gernaey KV, Eliasson Lantz A, and Larsen, A. 2009. Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. Biotechnology Progress 25: 1561-1581.

Abstract

Near-infrared (NIR) spectroscopy can potentially provide on-line information on substrate, biomass, product, and metabolite concentrations in fermentation processes, which could be useful for improved monitoring or control. However, several factors can negatively influence the quality of chemometric models built for interpretation of the spectra, thus impairing the analyte concentration predictions. The aim of this review was to provide an overview of necessary conditions and challenges that one has to face when developing a NIR application for monitoring of cell culture or fermentation processes. Important practical aspects are introduced, such as sampling, modeling of biomass concentration, influence of microorganism morphology on the spectra, effects of the hydrodynamic conditions in the fermenter, temperature influence, instrument settings, and signal optimization. Several examples from the literature are provided, which will hopefully guide the reader interested in the topic. Furthermore, the general procedure used for the development of calibration models is presented, and the influence of microorganism metabolism - potential source of correlation between analytes - is commented. Other important issues such as wavelength selection and evaluation of robustness are shortly introduced. Finally, some examples of potential applications of NIR monitoring are provided, including the implementation of control strategies, the combination with other monitoring tools (the so-called sensor fusion), and the description of process trajectories. On the basis of the review, we conclude that acceptance of NIR spectroscopy as a standard monitoring tool by the fermentation industry will necessitate considerably more on-line studies using industrially relevant - and highly challenging - fermentation conditions (high aeration intensity, high biomass concentration and viscosity, and filamentous production strain).

3.1 Introduction

Over the past decade, a great number of studies have explored the potential of near-infrared (NIR) spectroscopy for monitoring of bioprocesses, where it may provide real-time measurements of a range of different analytes. Such measurements can be useful in the frame of several applications, for example within automatic control, process optimization, or on-line quality assessment. However, the successful implementation of NIR spectroscopy for monitoring of bioprocesses is still not routine because the bioprocess poses several challenges.

An excellent review on NIR monitoring applied to bioprocesses was elaborated by Scarff et al. (2006). In this review, the evolution of NIR monitoring technology is presented based on the type of fermentation process that is run. Here an alternative application-oriented view will be given, focusing on the different practical aspects concerning the implementation of NIR for bioprocess monitoring, with particular emphasis on the on-line configuration.

First, a short introduction to the fundamental principles of NIR spectroscopy will be given. Next, the practical aspects when NIR monitoring is applied to fermentation processes will be discussed. Finally, some examples will be provided about the applications of NIR monitoring and some extensions of its use.

Table A.1 complements the text. It is an overview table that lists relevant details for all the literature references that have been consulted for the elaboration of this review. This table may be used for multiple purposes, e.g. consult wavelength regions used in the literature to model different analytes, compare results from different studies, find the relevant literature for a given type of fermentation process or NIR settings., etc.

3.2 Fundamentals of Near-Infrared Spectroscopy

3.2.1 Measurement principles

NIR spectroscopy is a spectroscopic method using the NIR region of the electromagnetic spectrum (from about 800 nm to 2500 nm). NIR spectroscopy focuses on vibrational energy changes of matter caused by rhythmic variations in the dipole moment of molecules (Willams and Norris, 1987).

Vibrational energy changes are related to frequencies of the electromagnetic spectrum lying within the mid-infrared (mid-IR) region between 2500 and 15000 nm (Willams and Norris, 1987). When molecules are irradiated, they can absorb photons that have an amount of energy coincident with the characteristic vibrations of the molecule; absorption then results in the excitation of the molecule to a higher energy level (Willams and Norris, 1987). Although fundamental absorption occurs within the mid-IR region of the spectrum, overtones and combination bands of fundamental absorptions manifest within the NIR region (Willams and Norris, 1987).

Photons with twice or three times the amount of energy necessary to elevate a molecule to the energy level corresponding to a fundamental absorption will cause excitation to the second or third energy levels, thereby creating a first and second overtone (Willams and Norris, 1987). However, few molecules manifest overtones, and the higher the overtone, the lower its probability of occurring. Therefore, the

first overtone band is generally much weaker in intensity than the fundamental absorption, and the second and higher bands are even weaker. First and second overtones occur at approximately one half and one third of the wavelength of the fundamental absorption, i.e. in the NIR region of the electromagnetic spectrum (Willams and Norris, 1987).

Combination bands occur when the absorbed photon excites two or more vibrations simultaneously. For this to happen, the energy of the photon has to equal the sum of the energies of the coupling vibrations. As a first approximation, the absorption intensities for combination bands are inversely proportional to the number of different vibrations that couple in the combination band (Willams and Norris, 1987).

When light passes through a material it may be selectively absorbed, while the remaining part of the light may be partially transmitted and reflected. The amount of light that is absorbed and the wavelength-dependent nature of absorption can provide information about the physical and chemical nature of the material. Unfortunately, the amount of absorbed light cannot be measured directly; instead, it has to be calculated from either the transmitted or the reflected light.

According to the Lambert-Beer law, the concentration of an absorber c is directly proportional to A, the sample absorbance, and inversely proportional to l, the optical path length:

$$c \propto \frac{A}{l} \tag{3.1}$$

The absorbance is defined as follows:

$$A = \log \frac{I_0}{I_t} \tag{3.2}$$

where I_0 is the intensity of the incident radiation, and I_t is the intensity of the transmitted radiation (Willams and Norris, 1987). The same equation may be used with reflection measurements, where the intensity of the transmitted radiation may be substituted by the intensity of reflected radiation. However, one must note that while the mechanical path length is fixed in transmission measurements (it is the length of a cuvette, or the slit of an in situ transmission probe), the optical path length depends on each sample.

As radiation passes through the sample, it crosses several interfaces between the actual material under study and the media surrounding it. At each interface, some of the radiation may be reflected. When the surface of the material studied is rough, some radiation penetrates into it, and reflection occurs at each microscopic interface, resulting in reflected radiation in all the directions. This type of reflection is known as diffuse reflection (as opposed to specular reflection, occurring only in one direction), and it may contain chemical information on the substance.

The amount of light that is reflected depends on the physical properties of the sample, such as the interfacial area. For example, powders have a large amount of interfaces, and thus they reflect most of the radiation that they receive. As the particle size decreases, more and more radiation is reflected (Willams and

Norris, 1987). Likewise, in a fermentation process, microorganisms with changing morphologies may be expected to reflect more radiation when for example fragmentation of pelleted biomass occurs.

In conclusion, NIR spectroscopy can be based on transmission measurements or on diffuse reflectance measurements, but both of them will be affected by scattering of light through the different interfaces. Scattering is viewed generally as a problem to solve through mathematical pretreatments of the spectra (e.g. Savitzky-Golay derivatives (Savitzky and Golay, 1964), multiplicative scatter correction (MSC) (Geladi et al., 1985), standard normal variate (Barnes et al., 1989), orthogonal signal correction (Wold et al., 1998)), although it may provide interesting information about the physical properties of the sample (Rodrigues et al., 2008).

3.2.2 Chemometrics

Chemometric techniques allow the spectroscopist to follow trends in absorbance spectra and explore the relationship with, e.g. analyte concentration changes. Multivariate analysis techniques, such as principal component analysis (PCA) and partial least squares (PLS), are commonly used to deconvolute and interpret spectroscopic data.

PCA is a method for re-expressing multivariate data. It allows to reorient the data, so that the first few dimensions account for as much as possible of the available information in the data set. If there is substantial redundancy present in the data set, then it may be possible to account for most of the information in the original data set with a relatively small number of dimensions or principal components (Lattin et al., 2003). The principal components are linear combinations of the original variables and have the property that each component is uncorrelated with the other principal components, which has the advantage of eliminating multicollinearity when using the PCA results in an analysis of dependence (e.g. regression analysis) (Lattin et al., 2003).

PLS regression - also known as projection to latent structures - is another type of multivariate analysis. It is a method to relate two data matrices, X (typically the spectroscopy data) and Y (typically the reference data), by a linear multivariate model, but goes beyond traditional regression in that it also models the structure of X and Y (Wold et al., 2001).

PLS shares an advantage with PCA, which is that the method can be applied for analysis of data sets containing strongly collinear variables, as often seen for fermentation process data. It can also model noisy and even incomplete variables in both X and Y. Moreover, it has the desirable property that the precision of the model parameters typically improves with an increasing number of relevant variables and observations (Wold et al., 2001).

3.3 NIR Spectroscopy Applied to Fermentation Processes

The application of NIR spectroscopy to monitor fermentation processes poses some challenges related to both the complex nature of the fermentation process

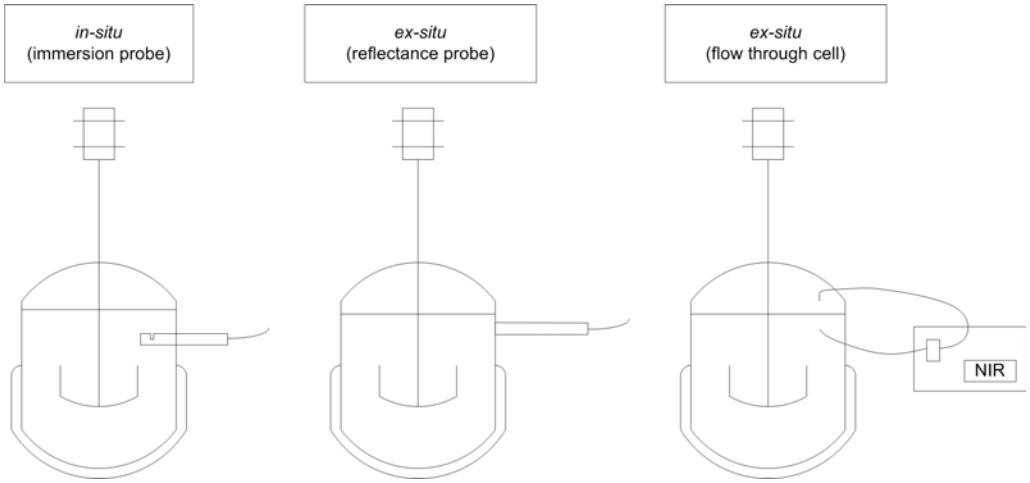


Figure 3.1: On-line sampling configurations. From left to right: in-situ measurements using an immersion probe, ex-situ measurements using a reflectance probe on the glass wall of the reactor, and ex-situ measurements using a flow-through cell (or loop).

itself (e.g. vigorous agitation and aeration, biomass growth, possible microorganism morphology changes, and correlation between analytes) and the technical difficulties of implementing NIR spectroscopy on-line. Therefore, a multidisciplinary approach is required to obtain a successful NIR monitoring solution.

3.3.1 Sampling

There are three main types of measurement approaches associated with NIR spectroscopy (Scarff et al., 2006):

- Off-line: a sample is taken and analyzed later, often at a different location (laboratory). Off-line transmission measurements with liquids usually use a cuvette of known path length and optical characteristics.
- At-line: a sample is analyzed immediately after being taken (rapid off-line).
- On-line: the process is analyzed directly and without manual sample handling (Figure 3.1). It can be divided into: (a) In situ (also known as in-line): the analysis device is a fiber-optic probe immersed into the fermentation broth, carrying the spectral information from the sample to the spectrometer. (b) Ex situ: The analysis device is physically outside the fermenter, and collecting measurement data usually involve either a flow-through cell or a fiber optic probe placed on the glass wall of the vessel.

In many production processes, conventional off-line analysis is performed daily or at the end of the fermentation, which results in a rather unreliable and inefficient analysis procedure (Arnold et al., 2002). At-line measurements offer an improvement to this situation: the analyzer is within the vicinity of the bioreactor and samples taken from the vessel are analyzed quickly, typically within

minutes (Arnold et al., 2002). However, the ideal approach is monitoring on-line, preferably in situ where the analyzer is in direct contact with the process stream, providing real-time determination of the analytes. One of the advantages of NIR spectroscopy is that NIR absorptions are generally 10-100 times weaker than the fundamental bands of mid-IR, enabling the direct analysis of samples without any sample preparation. Therefore, this technique is ideally suited for real-time measurements (Arnold et al., 2002).

Research in NIR spectroscopy applied to cell culture and fermentation has evolved from less challenging systems, with anaerobic conditions and/or low agitation, to more complex ones, characterized by vigorous agitation and aeration. Likewise, there has been a tendency to first implement at-line (Arnold et al., 2000, 2002; Riley et al., 2001; Vaidyanathan et al., 2000, 2001a,b,c) or ex situ systems (Ge et al., 1994; González-Vara et al., 2000; Hagman and Sivertsson, 1998; Vaccari et al., 1994), followed by the more challenging in situ (Arnold et al., 2003, 2002; Cimander et al., 2002; Cimander and Mandenius, 2002; Ferreira and Menezes, 2006; Navrátil et al., 2005; Rodrigues et al., 2008; Roychoudhury et al., 2007; Tamburini et al., 2003; Tosi et al., 2003), configurations when the technology and process knowledge became more mature.

Arnold et al. (2002) studied the differences between the at-line and the in situ approach for monitoring of biomass in an industrial *Escherichia coli* process. For the in situ configuration, first, the probe casing has to be able to withstand the pressure and temperature of sterilization. Second, depending on the process, the in situ collection of spectra is subject to disturbances from biomass particles, stirring, aeration, etc. (Arnold et al., 2002). Third, optimization of the signal is a crucial factor since the mode of spectral collection (transmission or reflectance, path length, etc.) cannot be changed without disrupting the sterile environment (Arnold et al., 2002). Finally, it was found that when using fiber optic cables the spectral region above 2100 nm is to a large extent unusable because of the fact that the fibers induce noise, which will have an adverse effect on the quality of the spectra in this wavelength region (Arnold et al., 2002). This upper region (the combination region) contains many important absorption bands for the analysis of biomass (Arnold et al., 2002; Vaidyanathan et al., 2001c, 1999), and different analytes such as glucose, ammonium, and lactate (Arnold et al., 2000, 2003; Finn et al., 2006; Lewis et al., 2000; McShane and Coté, 1998; Rhiel et al., 2002; Riley et al., 1997; Roychoudhury et al., 2007; Vaidyanathan et al., 2001a,b,c). In the study by Arnold et al. (2002), it was found that the biomass absorbed strongly in the region between 2260 and 2270 nm in the spectra collected at-line. When transferring the model from at-line to in situ, it was necessary to optimize the signal in the lower wavelength regions (first and second overtone regions) to obtain a satisfactory model.

Other studies of in situ applications of NIR have, however, shown that information can be extracted from the region more than 2100 nm (up to 2380 nm) to monitor different analytes (Lewis et al., 2000; Rodrigues et al., 2008; Roychoudhury et al., 2007). The problem of noise is not discussed in these articles, but in the case of Rodrigues et al. (2008), the problem may have been partly overcome by averaging a large number of scans (128) to maximize the signal to noise ratio.

Application of an external recurrent loop for the NIR measurement of glycerol, acetic acid, iso-butanoic acid, and volatile fatty acids in biogas production from waste was studied by Holm-Nielsen et al. (2008) Furthermore, Vaccari et al. (1994) and González-Vara et al. (2000) obtained NIR measurements through an external loop for the control of lactic acid production. The advantages of this configuration include separation from the chaotic conditions inside the bioreactor, increased flexibility in the sample analysis (the external recurrent loop may be attached to different bioreactors (Holm-Nielsen et al., 2008)), and, in the right setup, representative sampling of highly heterogeneous bio-slurries (Holm-Nielsen et al., 2008). Potential problems include dead zones, introduction of artifacts, as well as sterility issues due to the pumping out, the subsequent analysis, and the pumping back of a stream of fermentation broth into the fermentation vessel (Arnold et al., 2003). The risk for contamination is particularly unacceptable at production scale; however for biogas production, the sterility is not an issue.

The other ex situ alternative, placing the probe on the glass wall of the fermenter has been studied by Ge et al. (1994) and Cavinato et al. (1990) for the prediction of cell density and ethanol, respectively. The advantage of this method is that the probe is not in direct contact with the fluid and can be mounted on the fermenter after sterilization, which minimizes the requirements for the probe. The main problem with this approach is that the mode of data collection is limited to reflectance, because transmission measurements are not possible without direct contact with the fermentation broth.

NIR measurements are essentially sampling techniques, and as such they should be subject to the principles of the theory of sampling (TOS) (Gy, 1998). It has been stated that sampling is fundamentally a representative mass reduction (Gy, 2004a). The key concept 'representative' has been defined rigorously by Gy (1998) as a combination of both accuracy and reproducibility. The TOS provides a description of correct sampling methods and a quantification of the sampling errors to answer the question of how and how much sample should be taken (Gy, 2004a). In short, correct sampling is achieved when two conditions are fulfilled, namely that (a) all constituents making up the material to be sampled have an equal probability of being selected, and (b) the integrity of the selected constituents is duly respected (Gy, 1998).

The relevance of the TOS in bioprocess monitoring by NIR spectroscopy is obvious. Because chemometric models require a calibration process using reference data, correct sampling must be ensured during both the optical (i.e. NIR measurements) and the reference sampling. As the scale of the process to be monitored increases, the required mass reduction during sampling will be larger. Consequently, a thorough understanding of the TOS is especially important concerning industrial applications. Mixing in large fermenters, often with relatively high viscosities, is a critical issue. The time delay between the optical measurements and the analysis of the reference data may also be significant. In addition, many other factors such as temperature or ambient humidity are to be considered.

Taking into account the TOS, the representativity of ex situ NIR measurements may be questioned. For example, when an ex situ probe is placed on the glass wall of a fermenter, there could be a risk of biasing the biomass concentration, because the centrifugal forces generated with the agitation may concentrate more

particles near the fermenter wall (Ge et al., 1994). Contrary to a much extended misconception, this sampling bias would not be constant, as warned by Gy (1998). Similarly, the position of an immersion probe and the connection of an external loop may be critical. Unfortunately, there are not many references to the TOS in the literature on NIR spectroscopy for bioprocess monitoring (there are of course exceptions, such as Holm-Nielsen et al. (2008)), perhaps because it was developed originally for materials of mineral origin. Nevertheless, the theory is universally valid for all types of materials (Gy, 2004a), and the literature available about this subject is very extensive. The classic reference is Gy (1998), although a very accessible and comprehensive introduction to the theory can be found in the series Gy (2004a,b,c,d,e).

Importance of the production organism

The production organisms set the requirements for the chemical and physical environment in the bioreactor. Some call for anaerobic conditions with very low agitation; others, like mammalian cells, require aeration with low agitation, due to their sensitivity to shear forces; whereas the majority of microorganisms on industrial scale demand vigorous agitation and aeration. Furthermore, the biomass itself has a great influence on the matrix to be analyzed, because the biomass concentration and the specific morphology affect the viscosity and scattering properties of the broth. Consequently, each production organism, but also each different strain will influence the complexity of the monitoring problem to be solved.

Modeling of Biomass Concentration

The biomass concentration influences the scattering properties of the fermentation broth and thus the overall absorption. Several studies have shown that an increased biomass concentration increases the scattering of the fermentation broth thus decreasing the light permeability.

Ge et al. (1994) used short-wavelength NIR spectroscopy in the wavelength range 700-1100 nm to monitor the cell density in a fermentation process over a wide concentration range (1-60 g/L). All data were obtained from fed-batch fermentations of *Saccharomyces cerevisiae*, run in a 5-L reactor. The measurements were performed noninvasively through the glass walls of the vessel using a bifurcated fiber-optic bundle to both excite the sample and collect the diffusely backscattered radiation. It was observed that when the concentration of biomass increased (the concentration of scattering material increases), the relative amount of backscattered light reaching the detector also increased, causing a decrease in apparent absorbance (Ge et al., 1994). This decrease was monotonic but not linear (slope decreases).

For probes based on the transmission principle, the increased scattering leads to an increase in the overall absorption, causing baseline shifts (Crowley et al., 2005; Finn et al., 2006; Tosi et al., 2003). However, it is not convenient to use raw absorbance values to model biomass, because variations of scattering could lead to inaccurate predictions. Besides, the combination of physical and chemical effects on light can make the results difficult to interpret. It is, thus, common

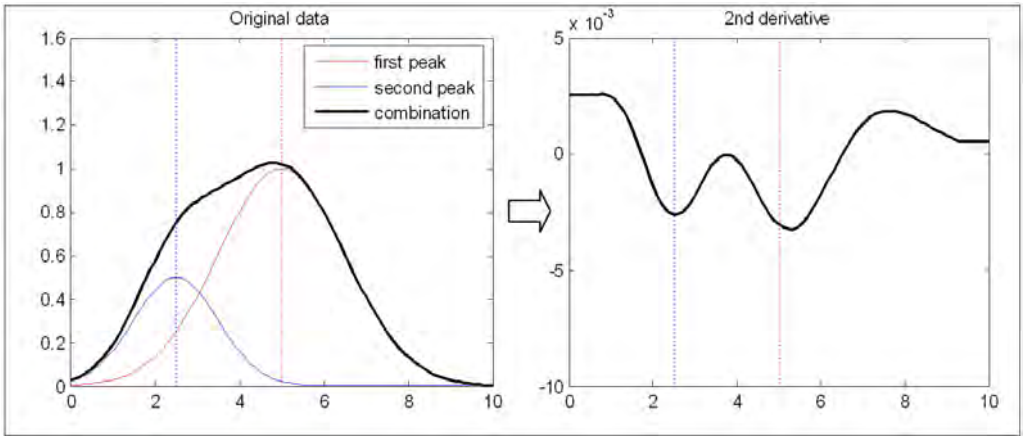


Figure 3.2: Second derivative can help resolve overlapping peaks. Note that it is not possible to determine the smallest peak (blue color) using the original data, whereas it is relatively easy after a pretreatment with second derivative. Note as well that peaks in the original data are observed as "troughs" after pretreatment with the second derivative method.

practice to pretreat the absorbance spectra before building a chemometric model. Second derivative is one of the most popular mathematical pretreatments, because besides removing baseline shifts, if the parameters are carefully chosen, it also aids to resolve spectral overlaps and show absorbance peaks more clearly (Figure 3.2)(Finn et al., 2006; Vaidyanathan et al., 2001b,c).

Vaidyanathan et al. (1999) studied the NIR spectrum of microbial biomass for a range of microorganisms: three of these were filamentous fungi (*Penicillium chrysogenum*, *Aspergillus niger*, and *Aureobasidium pullulans*), one a filamentous Gram positive bacterium (*Streptomyces fradiae*), and the other a Gram negative unicellular bacterium (*E. coli*). Three experiments were carried out. The first experiment was a comparative study of the NIR reflectance spectra of these organisms, and it showed that most of the prominent peaks occur at around the same wavelengths for the five microorganisms: for the combination band region, between 2270 and 2350 nm; for the overtone region, between 1650 and 1800 nm. In the second experiment, it was observed that the spectral signature of *P. chrysogenum* correlated well with dry cell mass measurements, when the biomass was isolated from its matrix and analyzed (Vaidyanathan et al., 1999). Finally, it was shown that the biomass reflectance spectral signature remained the same at different culture ages of this filamentous microorganism. The signal increased with culture time (because biomass was growing), but the prominent peaks remained in the same wavelength regions. Therefore, it was concluded that identified spectral regions may be useful for modeling of biomass regardless of the specific microorganism and the culture age. It was, however, also noted that these results do not imply that NIR spectroscopy may not be used to differentiate between microorganisms as there may be differences in the absolute or relative intensities.

Several studies (Arnold et al., 2002; Vaidyanathan et al., 2001a,b,c) have used

one of the microorganisms investigated by Vaidyanathan et al. (1999) and have, therefore, used the wavelength regions identified in this study to model biomass. Crowley et al. (2005) built a simple linear regression model for *Pichia pastoris* using the second derivative absorbance value at 944 nm. Finn et al. (2006) constructed a PLS model for *S. cerevisiae*, using the second derivative of absorbance spectra in the region 910-930 nm.

In a study by Tosi et al. (2003) NIR spectroscopy was used on-line to determine the biomass, glucose, lactic acid, and acetic acid during fermentations of *Staphylococcus xylosus* ES13. PLS regression was applied to spectra obtained in the 700-1800 nm region, yielding good results. Batch, repeated batch, and continuous fermentations were monitored and automatically controlled by interfacing the NIR sensor to the bioreactor control unit. A further aim of their investigation was to evaluate whether NIR was able to discriminate among morphologically different bacteria growing in the same defined medium under identical fermentation conditions. Comparison of spectra collected throughout fermentation runs of *S. xylosus*, *Lactobacillus fermentum*, and *Streptococcus thermophilus* showed the lack of any significant spectral difference specifically assignable to rod or spherical shapes. NIR was not able to discriminate between the three different strains because of the overwhelming presence of water in the spectra and the low final concentration reached by the biomass (<20 g/L).

Further work demonstrated the successful extension of a unique calibration model, developed for *S. xylosus*, to the other strains: *L. fermentum*, and *S. thermophilus*. The extension of the model to *S. thermophilus*, unlike *L. fermentum*, required two external validations, using a slight adjustment of bias and slope values of the curves, because *S. thermophilus* produces only lactic acid, and not acetic acid. All in all, the resulting model for *S. thermophilus* gave satisfactory results.

Although the aforementioned studies have shown that the wavelength regions at which different microorganisms absorb are essentially the same, and although in some cases it has even been possible to extend chemometric models from one microorganism to another (Tosi et al., 2003), it may still be possible to reveal the differences between microorganisms by means of a rigorous analysis of principal components (Vaidyanathan et al., 1999).

Decker et al. (2005) tested different preprocessing methods of NIR reflectance spectra combined with discriminant partial least squares regression (DPLSR) to classify different strains of *Penicillium camemberti* grown on cheese substrate. In DPLSR, the Y variables represent binary classification variables, with values 0 or 1. Among the different preprocessing methods, temperature- and wavelength-extended multiplicative scatter correction (TW-EMSC) was the only pretreatment that yielded a perfect classification of the strains using the lowest number of latent variables (LVs) (Decker et al., 2005). TW-EMSC removes undesired variations in baseline and signal amplification and extends the functionality of simple MSC with (a) corrections of linear and quadratic wavelength effects of, e.g. light scattering variations, and (b) elimination of patterns that look like temperature variation. To assess the relative importance of the different NIR wavelength regions in distinguishing the three isolates, the standardized regression coefficients of a DPLSR model were studied. The separation of one of the

strains from the other two was based on a large number of wavelengths, which could be due to differences in the degree of sporulation of the fungi (Decker et al., 2005). The other two strains had mirrored (symmetric) standardized regression coefficients, meaning that the important wavelengths to distinguish them were the same. According to the authors, this can be attributed to the ratio of the NIR absorbing compounds rather than differences in the chemical components (Decker et al., 2005).

A detailed determination of biomass composition or age has been typically done on dried or surface samples rather than submerged samples. In addition, although it is possible to separate biomass from off-line samples, it is more difficult to perform that operation on-line. All in all, it may be concluded that it is challenging to distinguish different microorganisms in submerged samples using exclusively the biomass NIR signal. Nevertheless, NIR spectra are influenced by every element of the whole matrix (or culture broth). Because every microorganism has its own characteristic metabolism and behavior, the matrix should help to identify abnormal situations such as a contamination. Trajectory plots, which will be presented later (Section 3.4.2), represent a good tool to identify these situations (Navrátil et al., 2005).

Influence of Morphology on the Spectra

Monitoring of fermentation processes using filamentous microorganisms is especially challenging. Besides requiring vigorous agitation and aeration rates, these microorganisms have a complex life cycle, involving filamentous growth and morphological differentiation (Borodina, 2007).

Vaidyanathan et al. (2003) have investigated the influence of morphology on the NIR spectra of the biomass of *S. fradiae*. The biomass was subjected to mechanical disaggregation to artificially generate the different morphological forms and computerized image analysis was used to characterize the morphological forms (unbranched, branched, single clumps, and clumps). The disaggregation treatment reduces the particle sizes and consequently increases the number of particles in suspension, which theoretically increases scattering (Willams and Norris, 1987); however, mathematical pretreatment of the spectra can minimize the scattering effects, facilitating the extraction of biomass specific spectral information. Vaidyanathan et al. (2003) applied a second derivative pretreatment on the NIR transmittance spectra of several mycelial biomass suspensions of different biomass concentrations and artificially induced morphologies. They observed that the effect of the disaggregation treatment seems to be that of a decrease of the biomass concentration (Vaidyanathan et al., 2003). According to the authors, this can be explained by the fact that for a unit volume, the incident radiation would be influenced more by clumps (denser biomass) than by dispersed morphology (less dense biomass) (Vaidyanathan et al., 2003).

The authors continued their work considering whether the sensitivity to morphology variations would affect the performance of chemometric models eventually built for a whole fermentation process. They calculated the correlation between biomass concentration and the spectral information in the wavelength regions related to biomass absorption and assessed to what extent changes in morphology impaired the correlations. Their study suggests that changes in morphol-

ogy influence the spectral variations to a lesser degree in the longer-wavelength NIR region (1600-2350 nm), compared with the visible or short-wavelength regions. Therefore, long-wavelength NIR spectral information is likely to be more useful in estimating biomass in mycelial bioprocesses (Vaidyanathan et al., 2003).

Taking the sensitivity to morphology as an advantage, NIR spectra may be useful in obtaining morphology related information. The NIR reflectance spectra of dried biomass were found to show correlations to the morphological variations introduced (Vaidyanathan et al., 2003). However, the reported correlation coefficients are not very high, which recommends interpreting the results with caution. As the authors state, further studies are required to assess the NIR spectra and their usefulness in providing morphology-related information (Vaidyanathan et al., 2003).

3.3.2 Conditions in the fermenter

Influence of the Hydrodynamic Conditions: Agitation and Aeration

The influence of the hydrodynamic conditions in the fermenter on NIR spectra obtained on-line has been studied by several authors (Ge et al., 1994; Navrátil et al., 2005; Tamburini et al., 2003). There is a consensus that variations in the hydrodynamic conditions in the fermenter change the spectra, and therefore during calibration and validation runs, the position of the probe in the fermenter, the agitation rate, and the aeration rate must be kept constant. However, the different observations made and the respective interpretations of this influence differ slightly as will be demonstrated below.

In the previously mentioned study by Ge et al. (1994), it was found that for a constant biomass concentration, a higher agitation rate caused an increase in the amount of backscattered light returned to the detector. The authors explained that this might be caused by an increased centrifugal force that impels the cells against the reactor wall, and/or by counteraction of the effect of settling of yeast particles. Conversely, at a constant biomass concentration, higher gas flow rates caused a decrease in the amount of backscattered light returned to the detector. According to the authors, with an increased number of air bubbles, a greater fraction of light was transmitted through the bubbles without scattering (Ge et al., 1994).

Tamburini et al. (2003) observed a similar influence of agitation rate, but an opposite influence of aeration rate. Note that they used an in situ transflection probe instead of the ex situ approach used by Ge et al. (1994) NIR spectroscopy was applied for on-line monitoring of glucose, lactic acid, acetic acid, and biomass in liquid cultures of the microorganisms *Lactobacillus* and *Staphylococcus*. A steam-sterilizable optical fiber probe immersed in the culture was used. This transflection probe had an optical slit of 1 mm, with an effective path length of 2 mm (Tamburini et al., 2003). A significant influence of the hydrodynamic conditions was observed. Both air flow rate and agitation caused baseline shifts (increasing the absorbance), which must be due to higher scattering effects of both particles and bubbles (Tamburini et al., 2003). However, the stirring rate had a greater influence, due to several factors: First, the movement of particles and bubbles caused more scattering; second, increasing the stirring rate created

more and smaller bubbles, which made it easier for them to flow through the slit of the probe. The effect of the air flow rate was less pronounced, because increasing the air flow rate increased the quantity of bubbles, but also enlarged them, making it more difficult for them to enter the slit of the probe (Tamburini et al., 2003).

In both articles (Ge et al., 1994; Tamburini et al., 2003), it is reported that increasing the agitation rate increases the scattering of NIR radiation. In the former case (Ge et al., 1994), this is translated in a greater diffuse reflectance signal, decreasing the apparent absorbance, whereas in the latter case (Tamburini et al., 2003), it means that less radiation is transmitted, in this case increasing the absorbance. Taking into account, the different measurement approaches (ex situ reflectance probe and in situ transflection probe), both conclusions are coherent: more agitation means more scattering.

However, comparing the observations on the effect of aeration rate, (Ge et al., 1994; Tamburini et al., 2003) the effect of bubbles in the spectra is unclear. Ge et al. (1994) reported that with more bubbles transmission of light became easier, whereas Tamburini et al. (2003) showed that more bubbles increased scattering. Perhaps, the explanation can be found in the different configurations used: ex situ probe vs. in situ transflection probe. It could be that the size of the bubbles in the former experiment were larger, facilitating the transmission, whereas in the latter experiment, the bubbles had to be smaller than 1 mm (recall the dimensions of the optical slit used) to enter the probe, therefore increasing scattering.

Navátil et al. (2005) presented similar observations to Tamburini et al. (2003). To compensate for the increase of apparent absorbance due to scattering effects of agitation and bubbles, they used a correction factor based on a fourth order polynomial equation, which was a function of the stirring rate. This correction improved the precision of biomass predictions (Navrátil et al., 2005). Their study will be presented with more detail in later sections (Section 3.4.1 and 3.4.2).

Temperature Influence

Temperature changes affect the vibration intensity of molecular bonds and therefore the NIR spectrum of a sample will be influenced by temperature variation (Cozzolino et al., 2007). In liquids containing water as its major component (as it is the case with fermentation broths), the effect of temperature changes is especially noticeable. Temperature affects the degree of hydrogen bonding and the hydration status associated with all constituents, and these changes influence the wavelengths at which overtones or combinations actually occur (Willams and Norris, 1987). An increase in temperature results in a decrease in the number of hydroxyl groups involved in hydrogen bonding, and consequently the absorption band of free hydroxyl increases (Cozzolino et al., 2007).

Water shows typical NIR absorption bands around 970 nm (O-H stretch second overtone), at 1450 nm (O-H stretch first overtone) and at 1900 nm (O-H stretching and bending vibrations) (Cozzolino et al., 2007). Hydrogen bonding lowers the frequency of water stretching vibrations while increasing the frequency of bending vibrations, therefore at higher temperatures, the stretching vibrations increase in frequency (Cozzolino et al., 2007).

Cozzolino et al. (2007) studied the NIR spectra of red and white wine samples

at six different temperatures. Between 1400 nm and 1500 nm, they observed a shift of the O-H stretch first overtone absorption band toward shorter wavelengths (higher frequency) with a temperature increase from room temperature (25 °C) to 50 °C. They also observed a similar shift of the O-H stretch second overtone absorption band.

There is not to our knowledge any specific work regarding the influence of temperature on NIR measurements applied to bioprocesses. However, Arnold et al. (2002) discussed the effect of temperature on their measurements as one of the challenges that may have to be solved when applying NIR monitoring in situ. In their case study, there was a temperature increase during the course of a fermentation run. The matrix was subjected to a sudden increase in temperature followed by a gradual increase. This resulted in highly complex spectral changes that were not readily predictable (Arnold et al., 2002). Unfortunately, in their process, the temperature increase is coupled with time and biomass growth, and, therefore, it is difficult to distinguish the effect of temperature from other effects. The authors state that once the spectra had been deconvoluted (using the second derivative) the analyte signals were more dominant compared to the effect of temperature change, and so it was still possible to determine analyte information from the spectra despite the temperature changes. Alternatively, the authors proposed to incorporate the temperature in the model as an independent variable, so that the model would account for the temperature change. However, no further details of how this could be implemented are given.

In the previously mentioned work from Decker et al. (2005), TW-EMSC (temperature- and wavelength-extended MSC) was used to reduce the influence of temperature on NIR spectra. The authors found that the water NIR signal variations may reflect differences in water binding properties between fungal strains. However, this water information may be difficult to use, because it is similar to the water variation caused by uncontrolled changes in sample temperature (Decker et al., 2005). In TW-EMSC, the spectral perturbations caused by temperature variations are summarized by PCA, and next the loadings of the principal components (two in total in this work) are included in the extended MSC regression model (Decker et al., 2005).

3.3.3 Instrument settings and signal optimization

Signal optimization prior to setting up a NIR monitoring system is crucial. As stated before, the in situ approach does not permit a change in the conditions for data collection without disrupting the sterile environment. Therefore, it is very important to get an optimal signal before beginning the data collection (Arnold et al., 2002).

Choice of Base System Technology

Selection of a suitable base technology is not trivial and may seriously affect the conclusions drawn from later spectral data. The generally accepted all-round technology is Fourier Transform Near-Infrared (FT-NIR), but spectrometers based on filters, diode arrays, mems chips, and dispersive gratings and similar have also been used successfully. All NIR spectrometers consist of a series of prin-

cipal components defining the usable wavelength range. The usable wavelength range of a system is defined by the response curves from vital components such as lamp, optics and detector. The base technology should provide a spectral response in a relevant spectral region, i.e. where molecules are expected to have a response, have a suitable net analyte signal to noise ratio, provide a suitable density of spectral information (resolution capability) and be stable over time. A very important difference between the technologies has been shown by Berntsson et al. (2001). Scanning grating systems have shown severe limitations when dealing with moving samples, which is very relevant to fermentation processes.

Choice of Probe

Selection of the most appropriate measurement principle (i.e. transmission or reflectance) is done according to the physical and chemical properties of the material to study. However, there are situations where these characteristics change with time. In a batch fermentation, for example, initial low biomass concentrations allow the use of transmitted light, whereas final high concentrations may require a reflectance measurement, because of the high number of particles in suspension. In these cases, a 'transflectance' measurement may result interesting, combining both transmission and reflectance measurements. Transflectance probes (for on-line measurements) have a window in contact with the fluid that is to be analyzed, separating the emitter or detector at one side, and a mirror, at the other side. Emitted light crosses the fluid, being partially absorbed, transmitted and reflected. The light that does not reach the mirror but, however, detected is due to reflection from the fluid, whereas the light that reaches the mirror and is reflected back to the detector constitutes the transmission component (Von Bargen, 1996). The combination of transmission and reflectance effects, despite increasing the versatility, may on the other hand make the results more difficult to interpret.

Optical Signal Optimization

With signal optimization we are referring here to the tuning of all those parameters which constitute the settings of the NIR equipment, either physical parameters such as the path length of a transmission probe or soft parameters, such as the number of scans, selection of the resolution in FT-NIR systems, etc. These values have an effect on the quality of the signal, they change the reproducibility or the signal to noise ratio. Changing the soft parameters is easier than varying the path length, but the effect of such changes is mainly observable after the data have been computerized. Therefore, a correct selection of all these values should be ensured before starting the monitoring. It is, however, difficult to find discussions on this matter in the literature. Most authors have their own preferred settings, but they do not give an explanation about why they chose those values.

The choice of path length of a transmittance or transflectance probe depends greatly on the physical properties of the fermentation broth. An opaque fluid, e.g. with a high biomass concentration, will require a short path length for sufficient light penetration whereas measurements in a clear solution can be made using a longer path length. Crowley et al. (2005) published an interesting study

concerning at-line monitoring of biomass, glycerol, methanol, and product in a high cell density *Pichia pastoris* fed batch process. It was found that a transmission probe with a path length of 0.5 mm could be used up to a certain biomass concentration (in their case study 64.1 g/L) above which the transmission signal was blocked. By switching the spectral collection mode to reflectance, it was possible to use NIR spectroscopy for monitoring of the process up to a biomass concentration above 110 g/L (Crowley et al., 2005). This is a type of temporally segmented modeling, which will be discussed further on (see Temporally Segmented Modeling Section).

According to the Lambert-Beer law, an increase in the path length will increase the absorbance signal. The dominating absorption peaks in the NIR spectrum of a fermentation broth are related with water absorption. The absorption at these locations will be several times higher than the absorbance in other parts of the spectrum. A long path length may be optimal for discerning absorbance bands in the region between 1600 and 1800 nm covering the first overtone of the Carbon-Hydrogen (CH) stretch but may lead to saturation around 1400 and 1900 nm (due to water absorption). Therefore, the choice of path length should take into consideration which part of the spectrum is expected to contain the absorption bands of the analytes of interest.

The optimal path length will also be dependent on the hydrodynamic conditions in the bioreactor. Arnold et al. (2002) investigated the effect of changing the path length between 0.5 and 2 mm on the quality of the spectra in an effort to optimize their signal. It was found that an increase in path length made the probe more susceptible to the intrusion of bubbles which interfered with the signal and reduced the reproducibility. In the majority of the published studies concerning in situ monitoring of fermentation processes using transmittance or transflectance probes, the path length is in the range between 0.5 mm to 2 mm (Arnold et al., 2002; Cimander and Mandenius, 2002; Navrátil et al., 2005; Rodrigues et al., 2008; Roychoudhury et al., 2007; Tamburini et al., 2003; Tosi et al., 2003).

Arnold et al. (2002) studied as well the effect of changing the number of scans. Triplicate measurements were taken comparing two values (32 and 64), and it was found that this number influenced only to a limited extent the repeatability. Therefore, it was chosen to use 32 scans, because the acquisition time is shorter. In other on-line applications, the reported number of scans is usually between 32 and 128 (Cimander and Mandenius, 2002; Navrátil et al., 2005; Rodrigues et al., 2008; Roychoudhury et al., 2007; Tamburini et al., 2003; Tosi et al., 2003).

In FT-NIR systems, the gains should in general be as high as possible, always avoiding saturation. The selection of the resolution is somewhat tricky. One would expect that a higher resolution (to avoid confusion, this means a smaller wavelength scanning window) will give more accurate spectra. However, as the resolution is increased, the signal to noise ratio is decreased, thereby reducing repeatability (Q-Interline unpublished data). If reproducibility is to be maintained, a higher number of scans will probably be required, thus increasing sampling time. Nevertheless, a very high resolution is generally unnecessary, taking into account that PLS and PCA methods are based on a linear combination of the absorbances at all the wavelengths: the information contained in bands is more important than the information contained in single wavelengths (particular

values are more attributable to noise rather than the real signal).

Influence of Probe Design. Multiplexing

The application of NIR spectroscopy in bioprocessing has been limited by its dependence on calibrations derived from a single bioreactor at a given time (Roychoudhury et al., 2007). In a recent article, Roychoudhury et al. (2007) proposed a multiplexed calibration technique, which allows calibrations to be built from multiple bioreactors run in parallel. This provides the flexibility to monitor multiple vessels and facilitates calibration model transfer between bioreactors. An FT-NIR system was used, with a multiplexer for up to six fiber optics channels. The measurements were taken on-line, using transflectance probes. PLS models were developed for glucose and lactate in a cultivation of Chinese Hamster Ovary cell lines. A slight decrease of the model quality was observed for the multiplexed models, in comparison with the conventional (single probe) models, and a simultaneous increase in the number of factors, as the model incorporates the interprobe variability.

The authors carried out signal intensity studies, assessing the factors that can have an effect on the spectra: spectrophotometer channels, probe design, and mirror optical properties. They identified the probe as the most influencing factor (Roychoudhury et al., 2007), hence pointing out the importance to ensure optical identity between probes in multiplexing applications. They continued with a probe test involving seven different probes, with a constant single spec channel and mirror, to capture only the variance contributed by the probes. Finally, they observed that more variance occurs at lower wave numbers, which seems logic because these regions are less energetic (Roychoudhury et al., 2007).

The Dominant Influence of Water

In dry mixtures, quantification of a given analyte is relatively easy, because its characteristic peaks in the spectrum allow the analyzer to recognize it. Straightforward correlations can be made between spectral information and analyte concentration. However, in solution (the fermentation medium is essentially a dilute aqueous solution), spectra of analytes are completely different from those of the anhydrous substances, with the characteristic peaks totally hidden by the overwhelmingly dominant presence of water (Tamburini et al., 2003).

Statistical techniques such as PLS or PCA provide a way to extract the information hidden under the spectra of aqueous mixtures. However, it would be preferable to have spectra accounting only for the influence of the analytes, eliminating the useless spectrum of water. One idea can be to use the spectrum of water or the medium used (e.g. serum, milk) as a reference for transmission measurements instead of taking air. Examples of this approach can be found in the literature (Cimander et al., 2002; Cimander and Mandenius, 2002; Navrátil et al., 2005). Unfortunately, the authors usually do not give an explanation why they decided to use this approach.

Tamburini et al. (2003) studied the possibility of subtracting the water spectrum from all the sample spectra. In their investigation, two calibration sets were built for each microorganism, one using difference spectra between samples and

water, and the other one using unprocessed NIR spectra. They stated that the first approach cannot be easily applied to the in situ, real-time monitoring of fermentations due to the large number of calculation steps required (Tamburini et al., 2003). Instead, water was considered as an additional component of the global matrix effect that influences the spectra. According to the authors, the drawback of such an approach is the need to increase the number of samples used to represent the system, because an additional variable has been introduced into the system (Tamburini et al., 2003).

The problem of subtracting the spectrum of water or the particular medium from the samples is that the 'background' is not static; it changes with the evolution of the fermentation. The influence of water will for example change with temperature and different salt concentrations. Therefore, it seems that subtracting the water spectrum at a given fixed set of conditions results in a baseline shift for each wavelength rather than a real elimination of the interference of water. Furthermore, subtracting the water spectrum may increase the noise and therefore wavelength selection becomes very important during the model-building phase.

3.3.4 Development of calibration models

In the 'standard' calibration procedure found in the literature (Figure 3.3) collected NIR measurements are correlated to off-line, reference analysis data of the studied matrix. The data set are usually divided into a calibration set, from which the standard error of calibration (SEC) or the root mean squared error of calibration (RMSEC) is obtained, and a validation set, giving a standard error of prediction (SEP) or a root mean squared error of prediction (RMSEP). This last value represents the robustness of the model developed (Arnold et al., 2002; Finn et al., 2006; Navrátil et al., 2005; Rodrigues et al., 2008; Tosi et al., 2003; Vaidyanathan et al., 2001c). The statistical techniques used in model development and validation are numerous. The number of LVs of a PLS model is usually optimized (Figure 3.4) by performing a cross-validation and minimizing the corresponding standard error of cross-validation (SECV) or the root mean squared error of cross-validation (RMSECV) (Navrátil et al., 2005; Rodrigues et al., 2008; Vaidyanathan et al., 2001c). Different pretreatments and wavelength interval selection methods are often tested (Navrátil et al., 2005; Rodrigues et al., 2008; Triadaphillou et al., 2007) to find a chemometric model as robust as possible (i.e. with the minimum SEP or RMSEP).

Besides a thorough understanding of chemometric techniques, a successful application of NIR spectroscopy for monitoring of bioprocesses requires knowledge of the dynamics of the fermentation process and its effects on the models. This section presents some advanced calibration strategies and briefly describes the application of wavelength selection methods and the assessment of the robustness of chemometric models as applied to fermentation processes.

Advanced Calibration Strategies

There are several types of challenges when developing a model, e.g. labor intense experimental work for data collection, correlation of spectra of different analytes

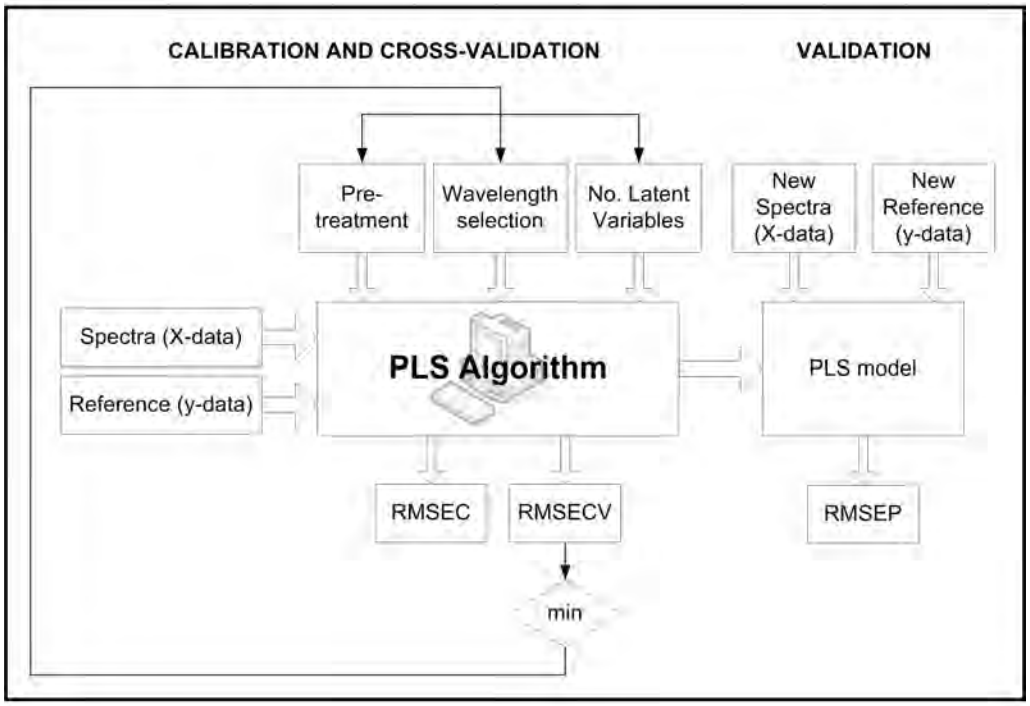


Figure 3.3: General calibration procedure, where the selection of pretreatment, wavelength intervals, and number of latent variables are driven by minimizing the RMSECV. The PLS model (and the settings selection) are used with new data to calculate the RMSEP, which indicates the robustness of the chemometric model. Note that this is a simple diagram, and there are of course variations of the optimization scheme.

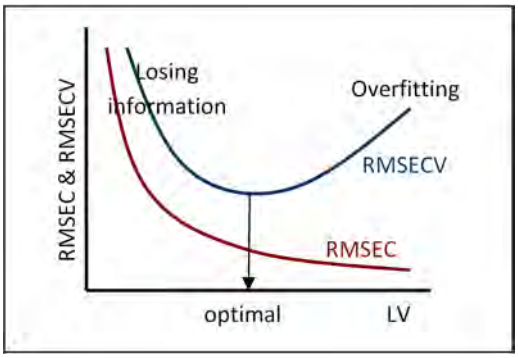


Figure 3.4: Optimization of the number of latent variables by minimizing the RMSECV.

in a matrix due to stoichiometric relations originating from metabolism, and too high biomass concentrations, which may result in a minimum in the transmission signal, and hence the reflectance mode may become a better strategy. Several calibration approaches have been developed to deal with these problems. They will be briefly presented and discussed below.

Rapid calibration

Riley et al. (1999) presented a new approach for generating NIR spectroscopic calibrations that requires significantly less time compared to standard calibration techniques. A small number of experimentally collected spectra of synthetic samples served as inputs to a computational procedure that simulated the collection of a larger data set by adding pseudo-random combinations of spectra for each component. The simulated spectra contained both analyte-specific and analyte-independent information. By using this strategy, it was possible to reduce the model development time from weeks to hours, maintaining the same error levels as obtained with full experimental calibrations (Riley et al., 1999).

If this method of model building proves successful in other systems, in terms of reproducibility and robustness, calibration and validation times can be dramatically reduced. This has an obvious industrial interest. However, extending this idea to an in situ application seems more difficult. One might indeed question whether it would be possible to first determine a small representative experimental data set and subsequently simulate all the variations that temperature changes, agitation and aeration (among other factors) can cause around these reference spectra.

Adaptive calibration approach

As mentioned earlier, the growth of microorganisms introduces correlations (intuitively, stoichiometric relationships) between the concentrations of nutrients, products, metabolites and biomass. This poses a problem when developing a calibration model as it is very often not known whether a model is reflecting variations of the corresponding analyte or whether it is reflecting variations of a different analyte which is correlated to the analyte of interest. For example, in a batch fermentation process where biomass is growing and a metabolite is concomitantly produced, the absorbance of the whole matrix will increase. Both models for biomass and metabolite will be based on this increase of absorbance, but are the models able to distinguish the metabolite produced and the biomass growth? As long as the yields of production and consumption of nutrients do not change between fermentation runs, this question seems irrelevant, but in any different situation, the models will produce wrong values. In other words, the robustness of the model will be compromised.

There are different opinions with respect to how a calibration model should be designed to deal with these natural correlations. Basically the literature offers three approaches: using synthetic mixtures, running real (representative) fermentations, and the intermediate situation, using semi-synthetic samples or 'adapted samples'.

Synthetic samples have been used in several studies for the development of cal-

ibration models. McShane and Coté (1998) calibrated PLS models for prediction of glucose, lactate, and ammonia in cell culture media based on a combination of synthetic samples and cell culture media samples. It was found that selectivity is produced by using data without correlations between the analytes of interest and specificity is obtained by including data from the sample type of interest. Similarly, Riley et al. (1998a) used a 'matrix-enhanced calibration procedure' for the calibration of PLS models for the prediction of glucose and glutamine in insect cell culture medium. This procedure uses a combination of synthetic samples and few media samples for the calibration of the prediction models.

Lewis et al. (2000) investigated the use of NIR spectroscopy for determination of glucose in cell culture media in situ in shake flasks. By using a combination of spectra collected from cell culture media and prepared media mixtures in the calibration of a PLS model, they eliminated the correlation between different analytes as well as the effect from the different background spectra in the different types of samples. Overall, there is agreement that a calibration based on purely synthetic samples yields poor results for prediction of media samples; however, addition of a few real media samples greatly improves the predictive abilities of the model for real media samples (Lewis et al., 2000; McShane and Coté, 1998; Riley et al., 1998a).

Riley et al. (2001) used a purely synthetic set of samples for the development of a NIR spectroscopy monitoring scheme for the quantification of 19 cellular nutrients and wastes in culture medium with and without serum mainly to avoid the influence of correlated analyte concentrations. The authors further argue that the use of synthetic samples provides accurate concentrations of many analytes over wide concentration ranges, and second, it provides a more severe test for the validity of the measurement methods (Riley et al., 2001). As an argument against this approach, one could mention the complete absence of cellular material in the system. However, the authors state that in hybridoma cultivations, the cell densities remain at a low enough level such that the impact of the cells and cell debris on FTIR measurements is negligible. Clearly, in fermentation systems with considerably higher cell density - this might be a problem.

Semi-synthetic samples were used by Riley et al. (1997) for the development of calibration models for the prediction of glucose and glutamine in insect culture medium. The semi-synthetic samples were generated by adding known and random amounts of glucose and glutamine to different aliquots of spent medium. However, in a subsequent paper by Riley et al. (1998b) it was found that this calibration strategy performed poorly on samples from additional bioreactor runs. Instead, it was found that reliable models could be calibrated by using an 'adaptive calibration scheme' also using semi-synthetic samples. In this approach, samples taken from the bioreactor at different time points were subsequently altered by adding different amounts of the analytes of interest.

Rhiel et al. (2002) successfully used the adaptive calibration approach to calibrate PLS models for prediction of glucose, lactate, glutamine, and ammonia in cell culture media based on NIR spectra collected off-line. They point out that one limitation of the adaptive calibration strategy is the slight distortion of the analyte distribution towards higher concentrations. The minimum concentration in the calibration samples may be higher than the minimum concentration in

the final application of the model; consequently, predictions will be based on extrapolation outside the calibration range (Rhiel et al., 2002). Nevertheless, this problem may partially be solved by combining dilution and spiking as described by Petersen et al. (2010) (Chapter 8). Finn et al. (2006) used adapted or semi-synthetic samples in addition to unaltered samples from a high cell density fed-batch baker's yeast bioprocess to successfully calibrate a model for prediction of glucose.

The use of real fermentation samples is defended by several authors (Arnold et al., 2003; Tamburini et al., 2003). Tamburini et al. (2003) applied NIR spectroscopy for on-line monitoring of glucose, lactic acid, acetic acid, and biomass in liquid cultures of the microorganisms *Lactobacillus* and *Staphylococcus*. The authors argue that it is not possible to ignore the interactions between the numerous components of the matrix, whose composition is almost impossible to predict (Tamburini et al., 2003). The composition of a fermentation fluid is the result of the instantaneous rates of substrate consumption and metabolite production, both influenced by environmental and physiological factors. Furthermore, the culture components are highly correlated, which makes it impossible to simulate fermentation broths with mixtures of pure compounds (Tamburini et al., 2003).

Arnold et al. (2003) argue that because of the nature of bioprocesses (everything is linked) an adaptive calibration approach should not be used. Instead, by adopting the traditional method of data collection, the authors had a greater confidence that the models would be transferable into an industrial setting, as the samples would not be altered or artificially generated in any way (Arnold et al., 2003). Furthermore, they add that as they were using an in situ configuration the adaptation of the samples was not practical. In their opinion, it would also defeat one of the main advantages of NIR spectroscopy, which is that no pretreatment of sample is required (Arnold et al., 2003). The last statement, although true for the calibration procedure, does not apply to the validation run or to the final application (industrial or not) of the model, where the samples will not have to be pretreated even though synthetic samples have been used in the calibration.

If it is not wanted or feasible to use synthetic or spiked samples in the calibration set another approach could be to build in variation via the experimental design, e.g. modifying the medium composition C/N ratio or C/P ratio. If including both 'normal' and 'unnormal' batches in the calibration set correlation between analytes might be reduced.

Temporally segmented modeling

Another limitation of standard linear calibration models (e.g. PCA and PLS) is their limited ability to accommodate variations in the matrix while maintaining a good accuracy. Although nonlinear models may improve the flexibility in processes where considerable changes in the matrix occur with time, one could a priori expect that dividing the process into phases based on matrix character might result in enhanced model quality using NIR spectroscopy (Arnold et al., 2001). That is the basic idea behind 'temporally segmented modeling.'

Arnold et al. (2001) used NIR to monitor an industrial bioprocess for the production of the antibiotic tylosin, using this temporally segmented modeling approach. Models were built over the entire time course of the fermentation (from

0 to 150 h), and also in two different phases: the growth phase (from 50 to 100 h) and the production phase (from 100 to 150 h). The SEP and the correlation coefficient show some improvements, which suggests that data segmentation is potentially a useful method to deal with high matrix variability (Arnold et al., 2001). The authors further argue that data segmentation may be useful in overcoming the limitations of the in situ NIR configuration which does not allow for changes in the probe path length or switching from transmission to reflectance with one probe (Arnold et al., 2001).

An important issue in the use of segmented models is how to cope with variations of the lag phase duration. Certainly, temporal segmentation will not always be as simple as switching from one model to another at a fixed time point (e.g. switching points at $t = 50$ and $t = 100$ h as suggested by Arnold et al. (2001)). Thus, it may be advantageous to use 'biological process time' instead of absolute time, leading to a new dynamic time scale reflecting the maturity state of the batch process (Jørgensen et al., 2004). The 'biological process time' is predicted from the on-line data. Once the analyte concentrations are determined by a model, these values can be compared with those obtained in a 'standard' process. In this way, it is possible to predict which relative time approximately corresponds to the state of the process (Jørgensen et al., 2004).

As it has already been mentioned, an alternative solution to accommodate profound changes in a fermentation process - however related to 'temporally segmented modeling' - is to change the mode of NIR spectral collection from transmission to reflectance. Crowley et al. (2005) observed that increases in the biomass concentration above a certain threshold value are not accompanied by proportional rises in the measured absorbances, and hence transmission measurements are of little value beyond this concentration limit. The authors advise that when using NIR spectroscopy it is important not to try to oversimplify the system by picking one mode of data collection and using it throughout the entire time course of the process. In their study, improved models could be built by switching from one mode of data collection to another mode (Crowley et al., 2005). However, it is important to note that with an in situ configuration this requires two probes, one for transmission mode and the other one for reflectance mode. Alternatively - as mentioned earlier - a transflection probe (Von Bargen, 1996) may be used, whose signal is combining both transmission and reflectance modes.

Wavelength Selection

PLS regression, as described in Section 3.2.2, is a multivariate analysis method allowing the extraction of information from a full spectrum and helping to deal with problems of collinearity. However, one question is still relevant: is it optimal to use information of the whole spectrum? Some wavelength regions of a spectrum correspond to absorbance bands of the analytes, yielding a high signal to noise ratio. On the other hand, there are other wavelength intervals where none of the analytes absorb. Unfortunately, the noise is still there, causing in this case a very poor signal to noise ratio. If a full spectrum is used, both signal and noise are modeled to a certain degree. The more variability the model should represent, the more PLS factors are required and the higher the risk for overfitting the

calibration data and hence reducing the model robustness (Vaidyanathan et al., 2001b). Several authors have shown that proper wavelength selection improves the model performance (Arnold et al., 2002; Navrátil et al., 2005; Rhiel et al., 2002; Triadaphillou et al., 2007; Vaidyanathan et al., 2001b).

Optimal selection of wavelength intervals may be achieved by use of criteria based on spectroscopy knowledge and statistics describing the performance of the calibration model (Bangalore et al., 1996). Numerous algorithms for automatic selection of optimal wavelength intervals have been developed. An extensive description of these methods is beyond the scope of this review. Only some examples of different wavelength interval selection strategies will be provided in this text, along with a very brief presentation of the most popular available algorithms, i.e. interval PLS (iPLS)(Nørgaard et al., 2000) and genetic algorithms (GAs) (Bangalore et al., 1996).

Arnold et al. (2002) identified the informative wavelength regions in their spectra in a qualitative way. By observing the spectrum of biomass from an at-line experiment, they could distinguish three main absorption bands. Later, they used one of the available bands in their in situ model. Alternatively, several authors have observed the spectra of the anhydrous analytes and have identified their absorption bands, which will be likely to give useful information in model building (Tamburini et al., 2003; Vaidyanathan et al., 1999). It is also common practice to study the spectra of aqueous solutions/suspensions of the major constituents of the fermentation broth in order to identify analyte absorption bands (Rhiel et al., 2002; Vaidyanathan et al., 2001a,b,c).

Yet, the earlier discussed strategies have some limitations and are perhaps more appropriate as a wavelength interval preselection method. The optimal wavelength intervals to build a model may not be the regions where the functional groups of a substance show the highest absorbance, because the signal of groups of different analytes may overlap (Triadaphillou et al., 2007). Despite the ability of PLS to resolve that information, still a better solution may be to use a region where the signal is not necessarily the highest, but instead more specific (Triadaphillou et al., 2007). As stated earlier, several computerized methods to select the optimal wavelength regions have been developed.

Rhiel et al. (2002) used a modified 'grid search' to optimize the spectral range. Basically, PLS calibration models were initially evaluated for a series of broad sections of the full range into smaller ranges with variable widths: 100, 200, 300, 400, and 500 cm^{-1}. The performance was judged on the basis of the prediction error. The best broad spectral range was then fine-tuned by stepwise expanding and contracting this range in a systematic fashion with sequential wave-number steps of 50, 30, 20 and 10 cm^{-1}.

Navrátil et al. (2005) applied the 'OPITIMIZM toolbox' to select spectral regions feasible for regression. For each iteration, a PLS regression of mean-centered values was run that tried to decrease the initial number of LVs and minimized the cross-validated variance of calibration. The wavelength elimination was based on the statistical test on heteroscedasticity, where the lowest statistical probability to accept elimination of wavelengths was set to 0.95. Based on the results from the signal selection analysis, a continuous spectral region was chosen for further work. Then, the number of PLS factors was optimized by using an external

validation.

Triadaphillou et al. (2007) have reviewed the area of wavelength selection. They divide the different strategies into three groups: dimension-wise selection, model-wise elimination, and subset selection. The authors focus on the 'subset selection' family of algorithms, whose distinctive feature is that they create different subsets of variables and the subset performance is evaluated. Algorithms that belong to this category include iPLS and GAs, which are the most widely applied methods (Triadaphillou et al., 2007).

iPLS (Nørgaard et al., 2000) is a methodology that develops local PLS models on spectral subintervals of equal width, and compares the prediction performance of these local models with the global (full-spectrum) model. The comparison is mainly based on the aforementioned parameter RMSECV, but other measures of the model quality may also be used. Leardi and Nørgaard (2004) comment that this technique is suboptimal since it is based on investigating a very small part of the very large solution space and that more exhaustive search methods such as GAs are more relevant. Yet, iPLS is still widely used (Rodrigues et al., 2008; Triadaphillou et al., 2007) as wavelength interval selection strategy due to its simplicity and lower computational burden compared to GAs (Triadaphillou et al., 2007). Moreover, iPLS is an attractive method in providing an overview of interesting spectral areas which could be selected. As the developers of iPLS state, the main contribution from using iPLS is the graphic output giving an overview of the spectral data (Nørgaard et al., 2000).

GAs (Bangalore et al., 1996) are optimization methods based on the principles of genetics and natural selection. A very comprehensive description of its application in NIR spectroscopy is given by Bangalore et al. (1996). GAs are powerful tools in variable selection, and after suitable modifications, they can also be powerful in detecting the most relevant spectral regions for multivariate calibration (Leardi and Nørgaard, 2004). However, a potential limitation of GAs is that this technique may generate models that are too specific to the model building data (in relation to wavelengths that contain apparent information by chance), which potentially may not predict future samples well (Triadaphillou et al., 2007). Furthermore, when spectral intensities are measured at a very large number of wavelengths the search domain increases correspondingly, and, therefore, the detection of the relevant regions is much more difficult (Leardi and Nørgaard, 2004). As a solution, Leardi and Nørgaard (2004) proposed the use of backward iPLS (biPLS) previous to the application of GAs. The operation of biPLS is similar to iPLS; however, in this case iPLS is applied to the data and then followed by backward elimination, each time eliminating the interval whose removal results in the lowest RMSECV (Leardi and Nørgaard, 2004). The procedure is either continued until the last interval or can be stopped when the number of retained wavelengths is lower than a predefined threshold, so that a preselected number of variables can be used as input for GAs.

A drawback with the mentioned wavelength selection methods is that the selected wavelengths are typically scattered throughout the spectrum. Triadaphillou et al. (2007) presented a new wavelength selection algorithm, Spectral Window Selection (SWS), which overcame this limitation. In this algorithm, the centers of the spectral windows and their widths are selected randomly and tested. Because

of the nonuniqueness of the windows selected when the algorithm is executed repeatedly, multiple models are constructed and these are then combined using stacking thereby increasing the robustness of the final calibration model. It was shown that SWS does improve predictions compared with models obtained from full spectrum analysis; however, SWS and GAs prediction errors were similar, and only slightly lower than iPLS errors (Triadaphillou et al., 2007).

Robustness: In-Batch and Batch-to-Batch Variability

Robustness is a very broad topic to tackle. Almost all the work that can be found in the literature assesses the robustness of the models. However, articles are typically focused on evaluating the performance of the models in a certain way: over a long period of time, on an industrial scale or forcing in-batch and batch-to-batch variability.

Vaidyanathan et al. (2000) explained that over an extended period of time, factors such as batch-to-batch variability in substrate feedstock, for instance, could contribute to spectral variations. The ability of the models to cope with such sample variations is critical (Vaidyanathan et al., 2000). NIR spectroscopy was used to predict substrate (oil) and product (tylosin) concentration in an industrial bioprocess that employed a heterogeneous medium for antibiotic production, using *S. fradiae* (Vaidyanathan et al., 2000). At-line measurements made by both the diffuse reflectance and transmittance modes were investigated, although diffuse reflectance measurements showed poorer results compared to transmittance measurements, especially for monitoring tylosin. An external validation was used, 2 years after the initial model development exercise, which demonstrates that chemometric models may be robust in performance over an extended period of time (Vaidyanathan et al., 2000).

Vaidyanathan et al. (2001c) assessed the robustness of models developed for the NIR spectroscopy prediction of mycelial biomass, total sugars, and ammonium, in a *P. chrysogenum* bioprocess, by challenging them with artificially introduced analyte and background matrix variations. The models were also challenged by using a data set from a process operated at a different scale from that used in the original model formulation (a smaller tank, with 1 impeller instead of 3, which probably introduced morphologic changes). The experiment was carried out over a period of 6 months, to randomize uncontrolled variations such as instrumental drift. Models developed for weak absorbers were vulnerable to changes in the matrix (Vaidyanathan et al., 2001c). A change in the scale of operation affected models influenced by biomass, probably due to changes of the morphology of the microorganisms (Vaidyanathan et al., 2001c). This investigation points out the importance of using an external validation set in order to identify useful models and anomalies within them (Vaidyanathan et al., 2001c).

In a recent article, Rodrigues et al. (2008) used NIR spectroscopy for simultaneous determination of multiple properties in a fermentation process using a filamentous microorganism: *Streptomyces clavuligerus*. The measurements were made in situ, by insertion of transflection probes into pilot and industrial bioreactors, and the authors succeeded in developing accurate NIR calibrations with a performance independent of scale and the specific bioreactor used (Rodrigues et al., 2008). This is specially challenging, since the process is run under vigorous

stirring and aeration rates, and the growth of filamentous bacteria as branched structures results in high viscosities in the liquid phase and non-Newtonian behavior. Furthermore, the rheological properties of culture liquids vary along cultivation time, from batch-to-batch and with reactor scale (Rodrigues et al., 2008). The data used as calibration in this work were acquired from 3 pilot bioreactors (geometric volume of 500 L) and three industrial reactors (20 m^3 and 40 m^3) (Rodrigues et al., 2008). The operating conditions were representative of those typically used in industrial processes for aerobic submerged cultivations at high cell densities, and the measurements were made under real industrial conditions such as high temperatures, humidity, and floor vibrations. The validation of the models was performed using data from a production plant bioreactor not used during the model development stage (Rodrigues et al., 2008). All in all, this work demonstrates the potential robustness of in situ NIR monitoring techniques applied to fermentation processes with filamentous microorganisms. It is a bit surprising that it was possible to build models for both N-source and C-source prediction using only 1 LV for each model (Rodrigues et al., 2008), especially considering the complexity of the matrix that is monitored and the numerous sources of variability between different experiments. Although a low number of LVs improves the robustness of the chemometric models, generally a larger number is required with complex matrices, because many different components and external factors introduce variability in the spectra. Some authors even comment that as a rule the number of LVs must exceed the number of components in the matrix (Tamburini et al., 2003). In Table A.1, it can be observed that the general trend in the literature is to use a moderate number of LVs.

3.4 Applications of NIR monitoring and extensions

NIR monitoring produces a large amount of data, which can be used for different purposes. If real-time data is available, a control unit can be interfaced with the monitoring equipment, potentially obtaining a fully automatic and selfregulating production system. In many processes, some substrates have an inhibiting effect depending on their concentration and the fermentation phase. This means that substrates have to be introduced in a fermentation process at the optimal concentration and at the optimal time. In this respect, NIR monitoring can for example be used to control the substrate feed addition, providing information both about starting point and rate, thus maximizing productivity. NIR monitoring can also assist in determining the optimal moment to harvest the batch. This is very important in yogurt production (Cimander et al., 2002; Navrátil et al., 2005), for example, or when degradation of the product takes place in the end of the cultivation, but there are many other examples.

NIR spectra can be used alone or the spectra can be combined with other complementary monitoring strategies. This last option allows supervision of process trajectories, assisting in the identification of abnormal situations such as a contamination. Besides, data can be used to gain understanding of the process. Kinetic models can be assessed by contrasting experimental data with theoretical predictions. The combination of soft and hard modeling has also been proposed ('soft' and 'hard' referring to empirical models and mechanistically derived mod-

els, respectively).

In this section, some of these applications and extensions of NIR monitoring will be briefly presented. Note that there are other applications of NIR monitoring that were not included here, and the text is meant to give an illustration of further use of NIR rather than providing a full coverage.

3.4.1 Implementation of control strategies

In González-Vara et al. (2000), a continuation of the work by Vaccari et al. (1994) a NIR-based control system allowed the full automation of a small-scale pilot plant for lactic acid production, using *Lactobacillus casei*. Three operation modes were tested: traditional continuous fermentation, continuous fermentation process with recycling of culture medium (MRF) and continuous fermentation process with cell recycling (CRF).

In the MRF operation mode, the fermentation broth was continuously fed back into the bioreactor after microfiltration and lactate extraction through ion-exchange columns, thus decreasing the lactate concentration in the culture broth in order to prevent its growth-inhibiting effect (González-Vara et al., 2000). In the CRF operation mode, the biomass was partially retained within the bioreactor to increase its concentration, thereby increasing glucose consumption. A high glucose concentration also has an inhibiting effect (González-Vara et al., 2000). Glucose, lactic acid, and biomass concentration were monitored on-line (trough a loop) by NIR spectroscopy. The data were used to control the flow of filtered culture broth in order to maintain either lactic acid (MRF mode) or glucose concentration (CRF mode) in the culture broth below inhibitory levels (González-Vara et al., 2000). In addition, the value of the biomass concentration was used to modulate the output of the controller to prevent biomass wash-out (González-Vara et al., 2000). Higher volumetric productivities were achieved in both operation modes compared to conventional continuous fermentation (González-Vara et al., 2000). The work by Tosi et al. (2003) has already been presented. NIR spectroscopy was used in situ to determine biomass, glucose, lactic acid, and acetic acid concentration during fermentations of *Staphylococcus xylosus* ES13. Batch, repeated batch, and continuous fermentations were monitored and automatically controlled by interfacing the NIR to the bioreactor control unit. Completely automated control was applied during continuous fermentations, where medium flow was modulated to maintain the glucose concentration at the desired level. Navrátil et al. (2005) applied in situ NIR spectroscopy and in-line electronic nose (EN) mapping to monitor and control a cholera-toxin producing *Vibrio cholerae* fed-batch cultivation. Prediction models for biomass, glucose, and acetate were developed using NIR spectroscopy, and growth rate control of the fed-batch cultivation based on biomass predictions was performed. In this process, acetate is produced under aerobic conditions in the presence of excess carbon source as a result of overflow metabolism. This overflow is undesirable, because it reduces growth rate, biomass yield, and recombinant protein production. A proportional-integral PI controller maintained the fed-batch process at the desired set point, minimizing acetate production and hence increasing protein yields.

3.4.2 Sensor fusion

Sensor fusion refers to the combination of NIR monitoring with other monitoring techniques such as ENs, mass spectrometry, and other standard instrumentation (pH, temperature, etc.).

ENs are arrays of chemical gas sensors, which in combination with NIR are very useful to generate process trajectories (Navrátil et al., 2005). In the study by Navrátil et al. (2005), the process trajectories were obtained by plotting the scores of the first principal component of the EN data set, versus the scores of the first principal component of the NIR data set. It can be observed that different processes of the same type follow the same pattern in these trajectory plots, within a certain interval. If the boundaries of a 'tolerance interval' are defined, it is possible to identify abnormal and unacceptable trajectories in future processes (Navrátil et al., 2005).

This was exemplified in Navrátil et al. (2005) by intentionally infecting the culture studied (*V. cholerae*) by an *E. coli* inoculum addition in the late batch phase. As the *E. coli* cells grew, the trajectory plot exhibited a larger and larger deviation from the normal behavior. Here, the importance of EN signals could be appreciated, because NIR signals of the contaminated fed-batch stayed more or less unaffected, whilst most of the variance was captured by EN signals, corroborating the fact that the EN is a powerful tool for detecting a microbial contamination.

Cimander and Mandenius (2002) applied for the first time multivariate statistical process control (MSPC) to analyze the data from a bioprocess on-line multianalyzer system consisting of an EN, a NIR spectroscope, a mass spectrometer and different probes providing standard bioreactor measurements (SBMs). The process under study was an *E. coli* fed-batch process for tryptophan production. A NIR immersion probe was used in the transmission mode.

Statistical process control (SPC) can be used to describe process abnormalities and deviations from standard target trajectories (Cimander and Mandenius, 2002). Four types of process trajectories may be calculated: average process trajectory, warning limit, action limit, and process experience limit (Cimander and Mandenius, 2002). MSPC is a combination of SPC with PLS modeling. Therefore, the authors used a similar strategy as with SPC, this time using only the first latent X variable for the modeling, which should have the highest correlation possible to the desired model output of biomass or tryptophan. With the LV selected, four latent process trajectories were calculated: average latent process trajectory, latent warning limit, latent action limit, and latent process experience limit.

It was observed that biomass MSPC models developed from the EN and SBM signals fell short, once new validation data exceeded the process experience limit (Cimander and Mandenius, 2002). On the other hand, it was detected that EN was more selective for tryptophan, whereas NIR and SBM models for tryptophan were based more on the measurement of biomass than on the measurement of tryptophan (Cimander and Mandenius, 2002). This is another good example of how different monitoring techniques can complement each other. In a different work, Cimander et al. (2002) used measurement data from an EN, a NIR spectrometer and standard bioreactor probes to follow the course of lab-scale yogurt

fermentation. It was desired to monitor the acidification of the culture causing the coagulation of milk caseins by the conversion of lactose to galactose and the formation of lactate leading to a decrease of the culture pH. The sensor signals were fused using a cascade neural network: a primary network predicted quantitative process variables, including lactose, galactose, and lactate; a secondary network predicted a qualitative process state variable describing critical process phases: inoculation, coagulation phase and harvesting. Navrátil et al. (2005) also applied NIR spectroscopy and EN data for on-line monitoring of yogurt and filmjölk (a Swedish yogurt-like sour milk) fermentations, but this time under industrial conditions (1000 L stainless steel tank in a dairy for regular production). First principal components for the NIR and the EN signals were used for on-line generation of a process trajectory plot visualizing the state of the fermentation (Navrátil et al., 2005). The NIR signals were also used to build PLS models for prediction of the culture pH and titratable acidity (Navrátil et al., 2005).

3.4.3 Combination of hard and soft modeling methods

Blanco et al. (2006) proposed the combination of hard and soft modeling methods to monitor alcoholic fermentations using the yeast *S. cerevisiae*. The work was divided into two parts. First, the authors tested various previously reported empirical hard modeling methods for the description of alcoholic fermentation. The product inhibition model of Hinshelwood (1946) was found as the best option. Second, fermentation processes conducted at variable temperature and pH were monitored on-line by using an immersion NIR probe. The results were processed by using multivariate curve resolution-alternating least-squares (MCR-ALS) in combination with the hard modeling information obtained in the first step as spectral equality constraints (Blanco et al., 2006).

Self-modeling curve resolution refers to a family of chemometric methods that factorize a mixed instrumental signal into the pure contributions associated to each component in a system (Blanco et al., 2006). These methods do not require reference analytical information, but are subject to inherent intensity and rotational ambiguities that entail imposing constraints to shorten the range of feasible solutions (Blanco et al., 2006). Even with such constraints, a band of feasible profiles rather than a unique solution has to be considered (Blanco et al., 2006). MCR-ALS can be used to minimize and optimize the range of feasible solutions by using a mixed approach introducing a hard modeling step based on a kinetic model as an additional constraint in the MCR-ALS algorithm (Blanco et al., 2006). A further explanation on these techniques is beyond the scope of this review and the interested reader is recommended to consult Blanco et al. (2006). This work confirms the efficiency of the joint use of soft and empirical hard modeling for studying evolving biological systems.

3.5 Conclusions

Throughout this text, multiple applications of NIR spectroscopy as a monitoring tool in fermentation processes have been presented. There is no doubt that there is a need for reliable and real-time measurements of key analytes involved

in bioprocesses. The direct measurement of biomass, substrate, and product concentration may help to gain understanding on the process, such that process optimization may be carried out and different control strategies may be proposed. Furthermore, on-line measurement of analytes enables real-time quality assessment. All these potential improvements with which NIR spectroscopy may contribute to the bioprocess are in accordance with the process analytical technology initiative launched by the Food and Drug Administration (FDA).

Also other promising bioprocess monitoring techniques have emerged the last decades involving other spectroscopy based technologies such as multi-wavelength fluorescence and dielectric spectroscopy (typically capacitance measurements), which have potential to provide important real time information (Kiviharju et al., 2008; Markx and Davey, 1999; Marose et al., 1999; Stärk et al., 2002; Yardley et al., 2000). Both methods have been successfully applied for determination of biomass concentration on-line (Boehl et al., 2003; Haack et al., 2004; Neves et al., 2000; Xiong et al., 2008); however, with multi-wavelength fluorescence it is also possible to monitor other compounds such as products (Clementschitsch et al., 2005; Haack et al., 2007; Lantz et al., 2006; Teixeira et al., 2009) under the condition that they are fluorescent. It is challenging to develop a spectroscopy monitoring application in bioprocesses. If real-time measurements are desired, the use of in situ probes may be required in processes where sterility is crucial. Immersing a probe in the fermentation broth implies that the sampling device will be subject to whatever the conditions present in the bioreactor. This may involve vigorous agitation and aeration and possible changes in temperature or pH. Furthermore, the measurement probes must be able to withstand sterilization treatments at high temperature and pressure. Besides, in batch or fed-batch processes, the measurements must be able to accommodate large variations in biomass concentration and thus for NIR different scattering conditions. Last but not least, in fermentation processes using filamentous microorganisms, chemometric models must tolerate morphological changes or different models for different phases have to be applied. This is especially significant since a large proportion of the antibiotics produced by the pharmaceutical industry are synthesized by filamentous microorganisms. Enzyme production by filamentous fungi is an important industrial application as well.

Vigorous agitation and aeration poses a larger challenge in NIR and fluorescence than for dielectric based spectroscopy. By probe development and the use of dual frequencies in dielectric measurements, the effect of gas bubbles can be eliminated (Kaiser et al., 2007; Kiviharju et al., 2008). Morphology changes affect the three methods in different ways. NIR is affected by morphology changes since it will affect the scattering properties. Biomass fluorescence is related to the content of aromatic amino acids, vitamins and proteins; hence, if a morphology change leads to a change in the biomass composition, the robustness of the models might be compromised. Change in the cell membrane composition and the shape or size of cells would influence in-line capacitance measurements as the capacitance is directly proportional to the cell volume. NIR has the advantage of being a broad method where several biological compounds can be measured, whereas the main drawbacks are the large water interference and the limited sensitivity. Fluorescence can detect compounds when present in very low concentrations,

but is restricted to fluorescent compounds. Dielectric spectroscopy is as already mentioned less sensitive to some of the disturbances in a bioprocess, but is only able to monitor biomass. However, with this method, it is the viable biomass that is determined and not the total. The methods are to some extent complementary and applying more than one at the same time may result in improved monitoring and control.

The scientific work in the last decade aiming at overcoming all the above mentioned challenges has been broad and extensive. Overall, it has been demonstrated that spectroscopy based methods in general, and NIR spectroscopy can be reliable and robust tools for monitoring of bioprocesses. Nevertheless, many of the publications which are available in the literature are related to at-line NIR measurements or on-line measurements using relatively simple fermentation matrices. For NIR spectroscopy to be accepted as a standard monitoring tool by the pharmaceutical and biotechnological industry, more on-line studies using highly challenging fermentation conditions are required that demonstrate reliability and simplicity. It is also crucial to demonstrate that NIR spectroscopy can provide higher value information compared to what traditional monitoring techniques can offer already, in an economically feasible way and using limited resources.

Bibliography

Arnold SA, Crowley J, Vaidyanathan S, Matheson L, Mohan P, Hall J, Harvey LM, McNeil B. 2000. At-line monitoring of a submerged filamentous bacterial cultivation using near-infrared spectroscopy. Enzyme and Microbial Technology 27:691–697.

Arnold SA, Crowley J, Woods N, Harvey LM, McNeil B. 2003. In-situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation. Biotechnology and Bioengineering 84:13.

Arnold SA, Gaensakoo R, Harvey LM, McNeil B. 2002. Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch *Escherichia coli* process. Biotechnology and Bioengineering 80:405–413.

Arnold SA, Matheson L, Harvey LM, McNeil B. 2001. Temporally segmented modelling: a route to improved bioprocess monitoring using near infrared spectroscopy? Biotechnology Letters 23:143–147.

Bangalore AS, Shaffer RE, Small GW, Arnold MA. 1996. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least-squares regression: Application to near-infrared spectroscopy. Analytical Chemistry 68:4200–4212.

Barnes RJ, Dhanoa MS, Lister SJ. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy 43:772–777.

Berntsson O, Danielsson LG, Folestad S. 2001. Characterization of diffuse reflectance fiber probe sampling on moving solids using a Fourier transform near-infrared spectrometer. Analytica Chimica Acta 431:125–131.

Blanco M, Peinado AC, Mas J. 2006. Monitoring alcoholic fermentation by joint use of soft and hard modelling methods. Analytica Chimica Acta 556:364–373.

Boehl D, Solle D, Hitzmann B, Scheper T. 2003. Chemometric modelling with two-dimensional fluorescence data for *Claviceps purpurea* bioprocess characterization. Journal of Biotechnology 105:179–188.

Borodina I, 2007. Functional genomics of *Streptomyces*. PhD Thesis at Center for Microbial Biotechnology (CMB), Department of Systems Biology.

Cavinato AG, Mayes DM, Ge Z, Callis JB. 1990. Noninvasive method for monitoring ethanol in fermentation processes using fiber-optic near-infrared spectroscopy. Analytical Chemistry 62:1977–1982.

Cimander C, Carlsson M, Mandenius CF. 2002. Sensor fusion for on-line monitoring of yoghurt fermentation. Journal of Biotechnology 99:237–248.

Cimander C, Mandenius CF. 2002. Online monitoring of a bioprocess based on a multi-analyser system and multivariate statistical process modelling. Journal of Chemical Technology and Biotechnology 77:1157–1168.

Clementschitsch F, Kern J, Pötschacher F, Bayer K. 2005. Sensor combination and chemometric modelling for improved process monitoring in recombinant *E. coli* fed-batch cultivations. Journal of Biotechnology 120:183–197.

Cozzolino D, Liu L, Cynkar WU, Dambergs RG, Janik L, Colby CB, Gishen M. 2007. Effect of temperature variation on the visible and near infrared spectra of wine and the consequences on the partial least square calibrations developed to measure chemical composition. Analytica Chimica Acta 588:224–230.

Crowley J, Arnold S, Wood N, Harvey L, McNeil B. 2005. Monitoring a high cell density recombinant *Pichia pastoris* fed-batch bioprocess using transmission and reflectance near infrared spectroscopy. Enzyme and Microbial Technology 36:621–628.

Decker M, Nielsen PV, Martens H. 2005. Near-infrared spectra of *Penicillium camemberti* strains separated by extended multiplicative signal correction improved prediction of physical and chemical variations. Applied Spectroscopy 59:56–68.

Ferreira AP, Menezes JC. 2006. Monitoring a complex medium fermentation with sample-sample two-dimensional FT-NIR correlation spectroscopy. Biotechnology Progress 22:866–872.

Finn B, Harvey LM, McNeil B. 2006. Near-infrared spectroscopic monitoring of biomass, glucose, ethanol and protein content in a high cell density baker's yeast fed-batch bioprocess. Yeast 23:507–517.

Ge Z, Cavinato A, Callis J. 1994. Noninvasive spectroscopy for monitoring cell density in a fermentation process. Analytical Chemistry 66:1354–1362.

Geladi P, MacDougall D, Martens H. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. Applied Spectroscopy 39:491–500.

González-Vara A, Vaccari G, Dosi E, Trilli A, Rossi M, Matteuzzi D. 2000. Enhanced production of L-(+)-lactic acid in chemostat by *Lactobacillus casei* DSM 20011 using ion-exchange resins and cross-flow filtration in a fully automated pilot plant controlled via NIR. Biotechnology and Bioengineering 67:147–156.

Gy P. 1998. Sampling for analytical purposes. Chichester, England: Wiley.

Gy P. 2004a. Sampling of discrete materials - a new introduction to the theory of sampling. I. Qualitative approach. Chemometrics and Intelligent Laboratory Systems 74:7–24.

Gy P. 2004b. Sampling of discrete materials. II. Quantitative approach - sampling of zero-dimensional objects. Chemometrics and Intelligent Laboratory Systems 74:25–38.

Gy P. 2004c. Sampling of discrete materials. III. Quantitative approach - sampling of one-dimensional objects. Chemometrics and Intelligent Laboratory Systems 74:39–47.

Gy P. 2004d. Part IV: 50 years of sampling theory - a personal history. Chemometrics and Intelligent Laboratory Systems 74:49–60.

Gy P. 2004e. Part V: Annotated literature compilation of Pierre Gy. Chemometrics and Intelligent Laboratory Systems 24:61–70.

Haack MB, Eliasson Lantz A, Mortensen PP, Olsson L. 2007. Chemometric analysis of in-line multi-wavelength fluorescence measurements obtained during cultivations with a lipase producing *Aspergillus oryzae* strain. Biotechnology and Bioengineering 96:904–913.

Haack MB, Eliasson Lantz A, Olsson L. 2004. On-line cell mass monitoring of *Saccharomyces cerevisiae* cultivations by multi-wavelength fluorescence. Journal of Biotechnology 114:199–208.

Hagman A, Sivertsson P. 1998. The use of NIR spectroscopy in monitoring and controlling bioprocesses. Process Control and Quality 11:125–128.

Hinshelwood CN. 1946. The chemical kinetics of the bacterial cells. London: Oxford University Press.

Holm-Nielsen JB, Lomborg CJ, Oleskowicz-Popiel P, Esbensen KH. 2008. On-line near infrared monitoring of glycerol-boosted anaerobic digestion processes: Evaluation of process analytical technologies. Biotechnology and Bioengineering 99:302.

Jørgensen P, Pedersen JG, Jensen EP, Esbensen KH. 2004. On-line batch fermentation process monitoring (NIR)-introducing 'biological process time'. Journal of Chemometrics 18:81.

Kaiser C, Carvell JP, Luttmann R. 2007. A sensitive, compact, in situ biomass measurement system controlling and monitoring microbial fermentations using radio-frequency impedance. BioProcess International 5:52–55.

Kiviharju K, Salonen K, Moilanen U, Eerikaeinen T. 2008. Biomass measurement online: the performance of in situ measurements and software sensors. Journal of Industrial Microbiology and Biotechnology 35:657–665.

Lantz AE, Jørgensen P, Poulsen E, Lindemann C, Olsson L. 2006. Determination of cell mass and polymyxin using multi-wavelength fluorescence. Journal of Biotechnology 121:544–554.

Lattin J, Carroll JD, Green PE. 2003. Analyzing Multivariate Data. Pacific Grove, CA, USA: Thomson.

Leardi R, Nørgaard L. 2004. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. Journal of Chemometrics 18:486–497.

Lewis CB, McNichols RJ, Gowda A, Coté GL. 2000. Investigation of near-infrared spectroscopy for periodic determination of glucose in cell culture media in situ. Applied Spectroscopy 54:1453.

Markx GH, Davey CL. 1999. The dielectric properties of biological cells at radiofrequencies: applications in biotechnology. Enzyme and Microbial Technology 25:161–171.

Marose S, Lindemann C, Ulber R, Scheper T. 1999. Optical sensor systems for bioprocess monitoring. Trends in Biotechnology 17:30–34.

McShane MJ, Coté GL. 1998. Near-infrared spectroscopy for determination of glucose, lactate, and ammonia in cell culture media. Applied Spectroscopy 52:1073.

Navrátil M, Norberg A, Lembrén L, Mandenius C. 2005. On-line multi-analyzer monitoring of biomass, glucose and acetate for growth rate control of a *Vibrio cholerae* fed-batch cultivation. Journal of Biotechnology 115:67–79.

Neves AA, Pereira DA, Vieira LM, Menezes JC. 2000. Real time monitoring biomass concentration in *Streptomyces clavuligerus* cultivations with industrial media using a capacitance probe. Journal of Biotechnology 84:45–52.

Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. 2000. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Applied Spectroscopy 54:413–419.

Petersen N, Ödman P, Cervera AE, Eliasson Lantz A, Stocks S, Gernaey KV. 2010. In situ near infrared spectroscopy for analyte-specific monitoring of glucose and ammonium in *Streptomyces coelicolor* fermentations. Biotechnology Progress 26:263–271.

Rhiel M, Cohen MB, Murhammer DW, Arnold MA. 2002. Nondestructive near-infrared spectroscopic measurement of multiple analytes in undiluted samples of serum-based cell culture media. Biotechnology and Bioengineering 77:73–82.

Riley MR, Amold MA, Murhammer DW. 1998a. Matrix-enhanced calibration procedure for multivariate calibration models with near-infrared spectra. Applied Spectroscopy 52:1339–1347.

Riley MR, Arnold MA, Murhammer DW, Walls EL, DelaCruz N. 1998b. Adaptive calibration scheme for quantification of nutrients and byproducts in insect cell bioreactors by near-infrared spectroscopy. Biotechnology Progress 14:527–533.

Riley MR, Crider HM, Nite ME, Garcia RA, Woo J, Wegge RM. 2001. Simultaneous measurement of 19 components in serum-containing animal cell culture media by fourier transform near-infrared spectroscopy. Biotechnology Progress 17:376–378.

Riley MR, Okeson CD, Frazier BL. 1999. Rapid calibration of near-infrared spectroscopic measurements of mammalian cell cultivations. Biotechnology Progress 15:1133–1141.

Riley MR, Rhiel M, Zhou X, Arnold MA, Murhammer DW. 1997. Simultaneous measurement of glucose and glutamine in insect cell culture media by near infrared spectroscopy. Biotechnology and Bioengineering 55:11.

Rodrigues LO, Vieira L, Cardoso JP, Menezes JC. 2008. The use of NIR as a multi-parametric in situ monitoring technique in filamentous fermentation systems. Talanta 75:1356–1361.

Roychoudhury P, Kennedy R, McNeil B, Harvey LM. 2007. Multiplexing fibre optic near infrared (NIR) spectroscopy as an emerging technology to monitor industrial bioprocesses. Analytica Chimica Acta 590:110–117.

Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36:1627–1639.

Scarff M, Arnold S, Harvey L, McNeil B. 2006. Near infrared spectroscopy for bioprocess monitoring and control: Current status and future trends. Critical Reviews in Biotechnology 26:17–39.

Stärk E, Hitzmann B, Schügerl K, Scheper T, Fuchs C, Köster D, Märkl H. 2002. In-situ-fluorescence-probes: a useful tool for non-invasive bioprocess monitoring. Advances in Biochemical Engineering/Biotechnology 74:21–38.

Tamburini E, Vaccari G, Tosi S, Trilli A. 2003. Near-infrared spectroscopy: a tool for monitoring submerged fermentation processes using an immersion optical-fiber probe. Applied Spectroscopy 57:132–138.

Teixeira A, Portugal C, Carinhas N, Dias J, Crespo J, Alves P, Carrondo M, Oliveira R. 2009. In situ 2D fluorometry and chemometric monitoring of mammalian cell cultures. Biotechnology and Bioengineering 102:1098–1106.

Tosi S, Rossi M, Tamburini E, Vaccari G, Amaretti A, Matteuzzi D. 2003. Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes. Biotechnology Progress 19:1816–1821.

Triadaphillou S, Martin E, Montague G, Norden A, Jeffkins P, Stimpson S. 2007. Fermentation process tracking through enhanced spectral calibration modeling. Biotechnology and Bioengineering 97:554–567.

Vaccari G, Dosi E, Campi AL, González-Vara A, Matteuzzi D, Mantovani G. 1994. A near-infrared spectroscopy technique for the control of fermentation processes: an application to lactic acid fermentation. Biotechnology and Bioengineering 43:913–917.

Vaidyanathan S, Arnold A, Matheson L, Mohan P, Macaloney G, McNeil B, Harvey LM. 2000. Critical evaluation of models developed for monitoring an industrial submerged bioprocess for antibiotic production using near-infrared spectroscopy. Biotechnology Progress 16:1098–1105.

Vaidyanathan S, Arnold SA, Matheson L, Mohan P, McNeil B, Harvey LM. 2001a. Assessment of near-infrared spectral information for rapid monitoring of bioprocess quality. Biotechnology and Bioengineering 74:376–388.

Vaidyanathan S, Harvey L, McNeil B. 2001b. Deconvolution of near-infrared spectral information for monitoring mycelial biomass and other key analytes in a submerged fungal bioprocess. Analytica Chimica Acta 428:41–59.

Vaidyanathan S, Macaloney G, Harvey L, McNeil B. 2001c. Assessment of the structure and predictive ability of models developed for monitoring key analytes in a submerged fungal bioprocess using near-infrared spectroscopy. Applied Spectroscopy 55:444–453.

Vaidyanathan S, McNeil B, Macaloney G. 1999. Fundamental investigations on the near-infrared spectra of microbial biomass as applicable to bioprocess monitoring. The Analyst 124:157–162.

Vaidyanathan S, White S, Harvey LM, McNeil B. 2003. Influence of morphology on the near-infrared spectra of mycelial biomass and its implications in bioprocess monitoring. Biotechnology and Bioengineering 82:715–724.

Von Bargen K. 1996. Transflectance probe having adjustable window gap adapted to measure viscous substances for spectrometric analysis and method of use. United States Patent 5,708,273, Appl No 08/647,24 Foss NIRSystems, Inc .

Willams P, Norris K, editors. 1987. Near-Infrared Technology in the Agricultural and Food Industries. St. Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Wold S, Antti H, Lindgren F, Ohman J. 1998. Orthogonal signal correction of near-infrared spectra. Chemometrics and Intelligent Laboratory Systems 44:175–185.

Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58:109–130.

Xiong ZQ, Guo MJ, Guo YX, Chu J, Zhuang YP, Zhang SL. 2008. Real-time viable-cell mass monitoring in high-cell-density fed-batch glutathione fermentation by *Saccharomyces cerevisiae* T65 in industrial complex medium. Journal of Bioscience and Bioengineering 105:409–413.

Yardley JE, Kell DB, Barrett J, Davey CL. 2000. On-line, real-time measurements of cellular biomass using dielectric spectroscopy. Biotechnology and Genetic Engineering Reviews 17:3–35.

Chapter 4

Multivariate data analysis

Many of the advanced sensors developed for bioprocess monitoring produce multivariate data sets, e.g. NIR and MWF spectroscopy. Often, the number of variables exceeds the number of observations and furthermore, there may be substantial collinearity in the data. This requires the use of multivariate statistical tools for visualization and identification of important correlations in the data. The most widely used methods for visualization and regression analysis of multivariate data in bioprocess monitoring is Principal Component Analysis (PCA) and Partial Least Squares regression (PLS) respectively. The two methods are presented below including important aspects such as pretreatment of data, validation, variable selection, and outlier detection.

4.1 Principal Component Analysis

Principal Component Analysis is an explorative tool which provides an overview of the data by representation in a new space spanned by the *principal components*. The principal components represent the directions of the largest variability in the data. Thus, the first principal component represents the direction of the largest variability, the second principal component represents the direction of the second largest variation and so forth for the higher order components. If there is considerable redundancy present in the data, it may be possible to account for most of the variation in the data using a relatively low number of principal components (Lattin et al., 2003). Algebraically, the principal components are orthogonal linear combinations of the original data with unit length. Using PCA, a data matrix $\mathbf{X}$, with I samples and J variables can be decomposed to

$$\mathbf{X} = \mathbf{1}\bar{\mathbf{x}}' + \mathbf{T}\mathbf{P}' + \mathbf{E} \tag{4.1}$$

Where $\mathbf{T}$ is the score matrix, $\mathbf{P}$ is the loading matrix and $\mathbf{E}$ is the residual matrix. The score matrix $\mathbf{T}$ represents the projections of the observations onto the principal components. The loading matrix $\mathbf{P}$ contains the coefficients in the linear combination of the original variables defining the principal components. The residual matrix $\mathbf{E}$ comprise the distances of the original variables to the projection on the principal components. PCA is widely used for visualization of multivariate data, outlier detection and identification of systematic patterns in the data (Martens and Martens, 2001). Furthermore, the results from the PCA (scores and loadings) can for example be used for for regression analysis (PCR), as input to neural networks or in statistical process control.

4.2 Partial Least Squares Regression

Partial Least Squares (PLS) regression is a tool for regression of dependent variable(s) (Y) on a multiple number of independent variables (X). Instead of estimating a regression coefficient for each of the independent variables, a new set of variables, the *latent variables*, are identified and used as regressors for $\mathbf{Y}$. With PLS regression information from both $\mathbf{X}$ and $\mathbf{Y}$ is used in order to estimate the latent variables (or components) and thus they capture the variation in $\mathbf{X}$ which is relevant for the description of $\mathbf{Y}$. If there is substantial redundancy in the data, the number of latent variables may be greatly reduced compared to the original variables. This rank reduction results in a more stable regression problem compared to the full multiple linear regression (Martens and Næs, 1989). The PLS calibration model can be summarized to (Martens and Martens, 2001):

$$\mathbf{T} = w(\mathbf{X}) \tag{4.2}$$

$$\mathbf{X} = \bar{\mathbf{x}} + \mathbf{T}\mathbf{P}' + \mathbf{E} \tag{4.3}$$

$$\mathbf{Y} = \bar{\mathbf{y}} + \mathbf{T}\mathbf{Q}' + \mathbf{F} \equiv \mathbf{b}_0 + \mathbf{X}\mathbf{B} + \mathbf{F} \tag{4.4}$$

where $w()$ is a function defined to maximize the covariance with the Y-variables. $\mathbf{T}$ is the score matrix which is used to model the X-data and the Y-data in terms of the X-loadings, $\mathbf{P}$, and the Y-loadings, $\mathbf{Q}$. $\mathbf{E}$ and $\mathbf{F}$ are the residual matrices. The Y-variables can be modeled directly from the X-variables via the regression coefficient matrix $\mathbf{B}$. A central points in the calibration of the PLS model is the determination of the number of components. It is often done based on evaluation of validation errors as described in Section 4.2.1 and 4.3. PLS regression is highly suitable for analysis of data with many collinear variables and a small number of observations, which is found in many practical calibration problems (Martens and Næs, 1989). Furthermore, it is relatively simple and easy to interpret the resulting models, which allows the user to gain insight into the underlying phenomena. This has made PLS regression one of the most widely used methods for analysis of spectroscopic and other data in bioprocess monitoring (Chapter 3, Lopes et al. (2004); Stärk et al. (2002)).

4.2.1 Number of components

One of the key points in the calibration of a PLS model (or other multivariate regression models) is the determination of the number of components. This should be chosen carefully in order to include all relevant information in the model but at the same time exclude noise, which is often dominant in the higher order components. The number of components is often chosen based on interval validation of the model, for example using cross-validation as described below. The errors calculated for the calibration set decrease continuously as the number of components increase. If a suitable validation strategy is chosen, the internal validation error will show a minimum or a knee point at the optimal number of components (Figure 3.4). This reflects the inclusion of important information in the first components and the presence of sample specific information (noise) in the higher order components.

4.3 Internal and External Validation

The multivariate statistical tools are very powerful and require careful validation. Validation is used in the development of the model, i.e. in the determination of the number of components, variable selection and evaluation of the data preprocessing. Furthermore, validation is used to asses the predictive ability of the calibrated model for future samples. Internal validation, such as cross-validation, is suitable in the development of the model. The predictive ability of the resulting model should preferably be evaluated based on external validation, i.e. validation on a new and independent data set. In cross-validation, the calibration and validation is repeated a number of times each time using part of the data for calibration and the remaining sample(s) for validation. The prediction error is then calculated based on an average of the prediction errors of the various validation samples. A few widely used cross-validation strategies are described below.

- **Leave-one-out cross-validation**: One sample is used for validation and the remaining samples are used for calibration. This is repeated for each sample.
- **Contiguous blocks**: The internal validation samples are selected as contiguous blocks of samples according to the order in which they appear in the data matrix. The number of blocks and iterations can be chosen by the analyst.
- **Random subsets**: The internal validation samples are chosen at random from the whole data set. The size of the validation set and the number of repetitions or iterations can be chosen by the analyst.
- **Selection of samples based on knowledge**: The subsets can be selected based on knowledge of the data, for example by defining the subsets such that all samples from a single experiment are left out of calibration and used for validation in each round. The procedure is then repeated such that all experiments are used for validation. This method is used in Chapter 8 and 9, and is referred to as leave-one-batch-out cross-validation.

Cross-validation and validation errors are often presented as the Root Mean Square Error (RMSE), which can be calculated as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{I}(\hat{y}_i - y_i)^2}{I}} \tag{4.5}$$

where $\hat{y}_i$ are the predicted values, y_i are the measured values and I is the number of samples.

4.4 Preprocessing

For some data types, preprocessing greatly improves the extraction of important information. Some of the most widely used preproccesing techniques are listed below. Mean centering and autoscaling are generally applicable, whereas the

derivative, multiplicative scatter correction and the standard normal variate are particularly suitable for correction of spectroscopic measurements.

- **Mean centering**: The mean for each variable is calculated based on all of the samples and subtracted from each measurement.
- **Autoscaling**: Mean centering and scaling of the data to unit variance. For each measurement, this is done by subtraction of the mean value and multiplication with the inverse standard deviation calculated for each variable based on all the samples.

$$x_{ij}^{Auto} = \frac{x_{ij} - \bar{x}_j}{s(\mathbf{x}_j)} \tag{4.6}$$

  where x_{ij} is the measurement for sample i variable j and $s()$ is the standard deviation.
- **Derivative**: The derivative is taken along the variable direction for each sample. The first and second derivative is often used for analysis of spectroscopic data because it can remove baseline effects and improve the spectral resolution (Figure 3.2). In order to reduce noise, it is often necessary to smoothen the data prior to calculating the derivative. This can be done with a Savitzky-Golay filter (Savitzky and Golay, 1964). The Savitzky-Golay filter smoothens the data by fitting a polynomial of a predefined order to a subset of the data points defined by the window width. This polynomial is then used in the calculation of the derivatives.
- **Multiplicative Scatter Correction (MSC)**: Each spectrum (or parts of it) is regressed against the mean spectrum and the slope (b) and off-set (a) are estimated (Geladi et al., 1985). These parameters are used to adjust each spectrum

$$x_{ij}^{MSC} = \frac{x_{ij} - \hat{a}_i}{\hat{b}_i} \tag{4.7}$$

  MSC thus adjusts each spectrum such that all appear to have the same level of scattering.
- **Standard Normal Variate (SNV)**: The mean and standard deviation is calculated for each sample spectrum and used to correct the individual spectra by (Barnes et al., 1989)

$$x_{ij}^{SNV} = \frac{x_{ij} - \bar{x}_i}{s(\mathbf{x}_i)} \tag{4.8}$$

4.5 Variable Selection

Most often some of the variables in a multivariable data set do not contain relevant information. The multivariate statistical tools described above are able to account for this to some extend. However, in many cases, predictions improve by removal of noisy variables which do not contribute with relevant information. A number

of tools are available for this such as genetic algorithms (Bangalore et al., 1996), iPLS (Nørgaard et al., 2000) or simulated annealing (Kirkpatrick et al., 1983). For a review readers are referred to Xiaobo et al. (2010). A genetic algorithm has been used in this work and is described in more detail below.

4.5.1 Genetic Algorithms

As the name implies, genetic algorithms (GA) is inspired by the selection process in natural evolution. The general procedure is outlined below (Lucasius and Kateman, 1991).

- **Initiation** Based on the complete data set a population of data sets with reduced variable sets are generated at random. Each population is evaluated based on a fitness criterion, for example the cross-validation error of a model calibrated based on the individual data sets.

- **Reproduction** The population reproduces by selection based on the fitness of each data set and by crossover and mutation. The best data sets have a higher probability of being selected for the new population. The crossover is a combination of two data sets at a random break point to produce two new data sets. In mutation, the variables are either added to or removed randomly from each data set at a given probability. The new population is evaluated based on the fitness criterion and the termination criterion. The procedure is repeated until the termination criterion is met.

- **Termination** The termination criterion is usually a convergence measure. For example the procedure can be stopped when 50% of the data sets in a population are identical. In addition, an upper limit of the number of generations can be set.

GA is a very flexible tool and may result in overfitting of the data, i.e. inclusion of noise in the model (Martens and Martens, 2001). This is particularly dangerous in data sets with a high number of variables and a low number of samples. Furthermore, for spectroscopic data, the results of GA selection may be difficult to interpret due to the selection of single wavelengths possibly spread over the entire spectrum. The algorithm can be adjusted to include groups of variables (or wavelength regions) instead of the single variables.

4.6 Outlier Removal

Outlying samples may have a large influence on the calibration of a model and may decrease its predictive ability. It is therefore beneficial to remove outliers prior to the calibration of a model. Basic tools for outlier detection in chemometric literature are based on the variation within the model space and/or outside the model space (Martens and Næs, 1989). Closely related measures such as the the Leverage, Hotelling's T^2 and Mahalanobis distance can be used to detect outliers inside the model space. The three measures are used in different chemometric software. The leverage of a sample (h_i) is a measure of the influence of the sample

on the predictions, i.e. the model, and can be defined as (Jobson, 1991)

$$h_i = 1/I + (\mathbf{x}_i - \bar{\mathbf{x}})' \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \tag{4.9}$$

where $\mathbf{X}^*$ represents mean deviation from $\mathbf{X}$ and $(\mathbf{x}_i - \bar{\mathbf{x}})'$ is the difference between the $i'th$ row of $\mathbf{X}$ and the mean $\bar{\mathbf{x}}$. $(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}$ thus acts as a weighting matrix. Similarly the Mahalanobis distance between a sample and the mean is given by

$$D^2 = (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{\Sigma}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) \tag{4.10}$$

where $\mathbf{\Sigma}$ is the covariance matrix of the data matrix $\mathbf{X}$. The central estimate of the covariance matrix is given by (Conradsen, 2003)

$$\hat{\mathbf{\Sigma}} = \frac{1}{I-1}\sum_{i=1}^{I}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \tag{4.11}$$

$$= \frac{1}{I-1}\mathbf{X}^{*\prime}\mathbf{X}^* \tag{4.12}$$

Thus the leverage is almost identical to the Mahalanobis distance but at a different scale. The Mahalanobis distance can for example be used to test if the two means ($\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$) of two normal J-variable populations with identical covariance structures ($\mathbf{\Sigma}$) but unknown parameters are identical based on I_1 and I_2 samples from the two populations respectively.

$$D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\hat{\mathbf{\Sigma}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \tag{4.13}$$

$$Z = \frac{I_1 + I_2 - J - 1}{J(I_1 + I_2 - 2)}\frac{I_1 I_2}{I_1 + I_2}D^2 \tag{4.14}$$

Z follows a $\mathrm{F}(J, I_1 + I_2 - J - 1)$ distribution if $\bar{\mathbf{x}}_1 = \bar{\mathbf{x}}_2$. This test is equivalent to the Hotelling's T^2-test. For a test of a single measurement belongs to the same population as the other measurement points, the calculations are reduced to

$$D^2 = (\mathbf{x} - \bar{\mathbf{x}})'\hat{\mathbf{\Sigma}}^{-1}(\mathbf{x} - \bar{\mathbf{x}}) \tag{4.15}$$

$$Z = \frac{I - J}{J(I-1)}\frac{I}{I+1}D^2 \tag{4.16}$$

which follows a $\mathrm{F}(J, I - J)$ distribution. The multivariate measures are powerful for identifying outliers because they use the information in the covariance structure of the data as illustrated in Figure 4.1. Figure 4.1 shows 500 random samples taken from a bivariate normal distribution with $\bar{\mathbf{x}} = [2,\ 3]$ and $\mathbf{\Sigma} = [1,\ 1.5; 1.5,\ 3]$. The black ellipse represents the 99% confidence interval based on Mahalanois distance (Equation 4.15 and 4.16). Using Mahalanonis distance it is possible to indentify the measurement point at [1.5, 6], marked by the red dot, as an outlier. This would not have been possible based on analyis of the singe variables individually.

For a bilinear model, such as a PLS model, the $\mathbf{x}_i$ and $\mathbf{X}$ can by replaced by the scores $\mathbf{t}_i$ and $\mathbf{T}$ to describe variation within the model space in equation 4.9 to 4.16. In the PLS toolbox (Eigenvector, WA, USA), Hotelling's T^2 is used as a

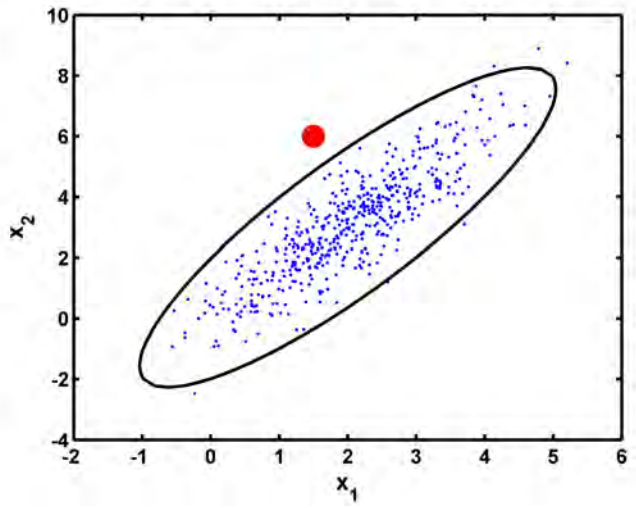


Figure 4.1: Illustration of the use of Mahalanobis distance for outlier detection. Blue dots: 500 samples taken from bi-variate normal distribution with $\bar{\mathbf{x}} = [2,\ 3]$ and $\mathbf{\Sigma} = [1,\ 1.5; 1.5,\ 3]$. Black ellipse: 99 % confidence interval. Red point: outlying sample.

measure of the variation of each sample within the models space of a multivariable model such as a PCA model.

The distance to the model space from the various samples can be quantified based on the residuals of both X- and Y-data. Analysis of the X-residuals can be used for both the calibration samples as well as new samples, whereas calculation of the Y-residuals requires knowledge of the measurement values of Y. For the X-data the residuals of a PLS model can be calculated as (Martens and Næs, 1989)

$$\hat{\mathbf{E}} = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}' - \hat{\mathbf{T}}\hat{\mathbf{P}}' \tag{4.17}$$

For each sample the X-residual can be written as

$$\hat{\mathbf{e}}_i' = \mathbf{x}_i' - \bar{\mathbf{x}}' - \hat{\mathbf{t}}_i'\hat{\mathbf{P}}' \tag{4.18}$$

For each element the X-residual can be written as

$$\hat{e}_{ij} = x_{ij} - \bar{x}_j - \sum_{a=1}^{A} \hat{t}_{ia}\hat{p}_{ja} \tag{4.19}$$

where A is the number of factors in the model. The Y-residuals can be calculated as

$$\hat{f}_{ik} = y_{ik} - \bar{y}_k - \sum_{a=1}^{A} \hat{t}_{ia}\hat{q}_{ka} \tag{4.20}$$

The residuals can be used to identify outlying samples or single outlying data points and they can be studied in more detail to provide information of the cause

of the deviating measurement. An effective way to analyze the residuals is by a simple graphical analysis (Martens and Martens, 2001). For a large number of samples and variables if may be beneficial to summarize the information for example in the residual standard deviation of individual calibration samples. For the X-residuals this is given by (Martens and Martens, 2001):

$$s(\hat{\mathbf{e}}_i)^2 = \frac{I}{\mathrm{df}} \sum_{j=1}^{J} \hat{e}_{i,j}^2 \tag{4.21}$$

where I is the number of calibration samples and df is the degrees of freedom, which can be approximated by (Martens and Martens, 2001)

$$\mathrm{df} = (IJ - J - A(\max(I, J))) \tag{4.22}$$

For evaluation of the $s(\hat{\mathbf{e}}_i)^2$ it can be useful to use an F-test approximation for the quantity

$$(s(\hat{\mathbf{e}}_i)/s(\hat{\mathbf{e}}))^2 \tag{4.23}$$

based on an F(df/I,df) distribution. The $s(\hat{\mathbf{e}})$ can be calculated as

$$s(\hat{\mathbf{e}})^2 = \frac{1}{\mathrm{df}} \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{e}_{ij}^2 \tag{4.24}$$

This test is a quick way to identify outliers in a large data set.

A range of commercial software are available for development of chemometric models. These include the Unscrambler (CAMO, Oslo, Norway), SIMCA (Umetrics, Umeå, Sweden) and the PLS toolbox (Eigenvector Research Inc, WA, USA) for MATLAB. Most programs include a wide variety of tools for the model development including the model types, preprocessing and variable selection algorithms and tools for outlier detection described above. Most of the programs are very intuitive and easy to use. However, it is important to have some knowledge of both the data and the chemometic tools in order to develop models that fit the intended purpose.

Bibliography

Bangalore AS, Shaffer RE, Small GW, Arnold MA. 1996. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least-squares regression: Application to near-infrared spectroscopy. Analytical Chemistry 68:4200–4212.

Barnes RJ, Dhanoa MS, Lister SJ. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy 43:772–777.

Conradsen K. 2003. En introduktion til statistik. Kgs. Lyngby, Denmark: IMM DTU, 6th edition.

Geladi P, MacDougall D, Martens H. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. Applied Spectroscopy 39:491–500.

Jobson JD. 1991. Applied Multivariate Data Analysis. Volume I: Regression and Experimental Design. New York, USA: Springer-Verlag, 621 p.

Kirkpatrick S, Gelatt CD, Vecchi MP. 1983. Optimization by simulated annealing. Science 220:671–680.

Lattin J, Carroll JD, Green PE. 2003. Analyzing Multivariate Data. Pacific Grove, CA, USA: Thomson.

Lopes JA, Costa PF, Alves TP, Menezes JC. 2004. Chemometrics in bioprocess engineering: process analytical technology (PAT) applications. Chemometrics and Intelligent Laboratory Systems 74:269–275.

Lucasius C, Kateman G. 1991. Genetic algorithms for large-scale optimization in chemometrics: an application. Trends in Analytical Chemistry 10:254–261.

Martens H, Martens M. 2001. Multivariate Analysis of Quality - An Introduction. Chichester, West Sussex, England: John Wiley and Sons Ltd.

Martens H, Næs T. 1989. Multivariate calibration. Chichester, West Sussex, England: John Wiley and Sons Ltd.

Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. 2000. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Applied Spectroscopy 54:413–419.

Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36:1627–1639.

Stärk E, Hitzmann B, Schügerl K, Scheper T, Fuchs C, Köster D, Märkl H. 2002. In-situ-fluorescence-probes: a useful tool for non-invasive bioprocess monitoring. Advances in Biochemical Engineering/Biotechnology 74:21–38.

Xiaobo Z, Jiewen Z, Povey MJ, Holmes M, Hanpin M. 2010. Variables selection methods in near-infrared spectroscopy. Analytica Chimica Acta 667:14 – 32.

Chapter 5

Experimental work

This chapter presents an overview of the various setups used in the experimental work and thereby provides a more detailed description of the data, which forms the basis of the studies presented in Chapters 6 to 9. The overall aim of the work was to study different sensors for monitoring of microbial cultivations. It was chosen to focus on filamentous cultivation systems, since they are of high industrial importance and at the same time represent particularly challenging systems for sensor applications. Furthermore, the majority of studies of advanced sensor technologies have so far focused on unicellular systems (Chapter 2 and 3).

The filamentous morphology directly influences the rheology of the cultivation broth. In submerged industrial cultivations with a high concentration of filamentous biomass, the broth is often highly viscous, which has a severe impact on the mixing and mass transfer. Furthermore, the morphology has been found to be connected to productivity in some systems (Pinto et al., 2004). Thus in filamentous systems, the morphology makes up an additional process variable which which should be monitored and controlled. Moreover, the morphology complicates other measurements in the broth. For example, the morphology may influence spectroscopic measurements, and has a severe impact on some of the traditional sensors for biomass measurements.

In this thesis, the monitoring of filamentous cultivations was studied from three different angles: (1) Quantification of filamentous morphology using laser diffraction and relation to the rheological parameters described in Chapter 6 and 7, (2) in situ NIR spectroscopy for quantification of substrates described in Chapter 8 and (3) evaluation of on-line biomass sensors described in Chapter 9.

Laser diffraction was used to quantify the morphology in *Streptomyces coelicolor* and *Aspergillus oryzae* cultivations. Both systems exhibit pelleted growth suitable for quantification with laser diffraction. The cultivations were designed to provide large variation in the morphology within each batch as well as between the different batches.

The on-line sensors, NIR and the various biomass sensors, where tested in *S. coelicolor* cultivations. Two approaches were taken to ensure sufficient variation in the data and to break the correlations in the data: (1) variation introduced by the use of semi-synthetic samples combined with standard cultivation samples and (2) variation introduced in the cultivation protocol. *S. coelicolor* batch cultivations combined with semi-synthetic samples were used for the study of in situ NIR spectroscopy presented in Chapter 8 and *S. coelicolor* fed-batch cultivations run as part of a process development were used for evaluation of the on-line biomass sensors presented in Chapter 9.

This chapter is structured as follows. First, the two model systems *S. coeli-*

color and *A. oryzae* are described. Following this, an overview is presented of the experimental design of the *A. oryzae* fed-batch and the *S. coelicolor* batch and fed-batch cultivations. In addition, this chapter provides a more detailed description of the observed growth and metabolism in the *S. coelicolor* cultivations.

5.1 Model systems

5.1.1 *Aspergillus oryzae*

Aspergillus oryzae is a filamentous fungi which belongs to the group of Fungi Imperfecti or Deuteromycetes, due to its asexual life cycle (Chang and Ehrlich, 2010). It has a long history of use in the production of several oriental foods such as soy sauce, sake, bean curd seasoning and vinegar production (Machida et al., 2008). *A. oryzae* is able to produce and secrete a large number of enzymes and in large quantities (Chang and Ehrlich, 2010; Machida et al., 2008). It is therefore widely used industrially for the production of enzymes, such as glucoamylase, alpha-amylases and proteases used for starch processing, baking, and brewing worldwide (Chang and Ehrlich, 2010; Machida et al., 2008)). Furthermore, it is widely used as a host for production of heterologous enzymes (Chang and Ehrlich, 2010; Christensen et al., 1988; Huge-Jensen et al., 1989). Due to its long history of use, it has earned the GRAS status (Generally Regarded As Safe) with the Food and Drug Administration in the United States of America, although it is closely related to aflatoxin producing species such as *A. flavus* (Chang and Ehrlich, 2010; Kobayashi et al., 2007).

The asexual life cycle of *A. oryzae* consists of germination of the spores (conidia), development of vegetative mycelia, and conidiation (Madigan et al., 2003). The vegetative mycelia are made up of single filaments termed hyphae, which grow by tip extension and branching (Madigan et al., 2003; Pedersen, 1999). In submerged cultures, the fungi may grow as either freely dispersed hyphal elements or as pellets, which are spherical agglomerates of several hyphal elements (Carlsen et al., 1996).

The complex morphology has large implications for industrial production. Dispersed growth may result in highly viscous broth with non-Newtonian behavior, which may cause problems for gas-liquid mass transfer and mixing (Carlsen et al., 1996). On the other hand pellet growth may result in mass transfer problems in the core of the pellet. A number of different process parameters have been shown to influence the morphology of *A. oryzae* and other filamentous organisms. These include agitation intensity (Amanullah et al., 2002), feed strategy (Bhargava et al., 2005), medium composition, pH and the inoculum type and concentration (Carlsen et al., 1996; Prosser and Tough, 1991).

5.1.2 *Streptomyces coelicolor*

The streptomycetes are gram-positive filamentous bacteria that have attracted much attention due to their extensive production of secondary metabolites. The streptomycetes are known to produce over 500 distinct antibiotic substances of which over 60 have found practical application in human and veterinary medicine, agriculture, and industry (Madigan et al., 2003). Furthermore, the stretomycetes

are characterized by a complex life cycle which involves morphological differentiation. *S. coelicolor* is the most genetically studied *Streptomyces* strain and it has served as a model organism for the streptomycetes for decades (Hopwood, 1999). In 2002 the complete genome sequence was released, which has further expanded the knowledge of the organism (Bentley et al., 2002). *S. coelicolor* produces a range of secondary metabolites including the two pigmented antibiotics: undecylprodigiosin and actinorhodin (Bibb, 1996).

The streptomycetes form branching filaments usually between 0.5-1.0 μm in diameter and of indefinite length (Madigan et al., 2003). Growth occurs at the tips of the filaments and is often accompanied by branching, resulting in a tightly woven network of filaments, i.e. the mycelium (Madigan et al., 2003). Grown on a solid surface, an aging colony forms sporophores, which are multinucleate aerial filaments projecting above the surface of the colony. The *Streptomyces* spores (conidia), are formed by formation of cross-walls in the sporophores followed by a separation of the individual cells directly into spores (Madigan et al., 2003). The formation of spores usually only take place when the organism is growing on a solid surface such as agar. In submerged cultures, the observed morphology ranges from dispersed mycelia to large dense pellets (Hobbs et al., 1989; Pinto et al., 2004). The morphology is dependent on a number of factors such as the inoculum type and concentration (Kieser, 2000; Manteca et al., 2008; Pinto et al., 2004), medium composition (Hobbs et al., 1989; Kieser, 2000) and mechanical stress (Kieser, 2000; Pinto et al., 2004), but only little is known about the mechanisms and overall kinetics that control the morphology (Pinto et al., 2004).

The secondary metabolism of the streptomycetes is growth-phase dependent and is usually associated with restriction of growth (Bibb, 1996; Doull and Vining, 1990). The supply of nutrients has a large effect on the onset and intensity of secondary metabolism. Limiting the availability of different nutrients may have different metabolic and regulatory effects (Doull and Vining, 1990). For *S. coelicolor* it has been observed that both nitrogen and phosphate has a suppressive effect on the secondary metabolism (Doull and Vining, 1990). Moreover, iron limitation has been found to enhance the intracellular production and export of actinorhodin. Finally, the pH affects both production and excretion of antibiotics in *S. coelicolor* (Elibol, 2002; Kim et al., 2007; Wright and Hopwood, 1976).

5.2 *Aspergillus oryzae* cultivations

Samples from *A. oryzae* cultivations were used to study the size distribution measured using laser diffraction and the correlation to the rheological properties. For this purpose, it was important to cover a wide range of both rheological properties and fungal morphologies. The cultivations were run as part of a normal pilot plant process development at Novozymes, and thus the main aim was fine tuning of the process prior to up-scaling. Variation in rheology and morphology were introduced by application of different feed strategies.

In total twelve fed-batch cultivations of *A. oryzae* were conducted in cultivation vessels of 550 liters total volume at Novozymes Fermentation Pilot Plant by Stuart Stocks. Different feed strategies were applied which could be described to various degrees as “aggressive” or ’conservative’ with a feed mode that can be

described as "pulse-pause" (constant feed rate interrupted by pauses of approximately 7 minutes) or "continuous" (uninterrupted feeding where the feed rate can be changed continuously). The pulse-pause effect is a patent protected method known to reduce the viscosity of the broth through controlled fragmentation or morphological control of the biomass (World patent: WO 2003/029439; Bhargava et al. (2005)). Off-line samples were collected and characterized with respect to rheology, biomass concentration and particle size distribution. The results confirmed that the use of different feeding strategies resulted in a large variation in broth viscosity and fungal morphology. The data set was thus suitable for exploring the relationship between viscosity and fungal morphology characterized by the particle size distribution. The results are presented in Chapter 7. Due to confidentiality reasons, no additional information of the cultivations can be given.

5.3 *Streptomyces coelicolor* batch cultivations

Streptomyces coelicolor batch cultivations combined with semi-synthetic samples were used for evaluation of in situ NIR spectroscopy for measurement of substrate concentrations (glucose and ammonium) described in Chapter 8. The main aim of the experimental work was to collect in situ NIR spectra and corresponding off-line samples in a challenging filamentous system covering a relatively wide range of concentrations. Furthermore, the substrate and biomass concentrations should be uncorrelated or at least only weakly correlated, in order to ensure analyte specific models. It was chosen to focus on two important substrates: glucose and ammonium, since they can be measured using NIR spectroscopy based on absorptions from the C-H, O-H, C-O and N-H bonds at the concentration levels in the broth. In the batch cultivations, strong correlations were observed between glucose, ammonium and biomass concentrations. In order to break the correlations, it was chosen to combine the data collected in standard batches with measurements of semi-synthetic samples spiked with various amounts of biomass, glucose and ammonium.

Nine *S. coelicolor* batch cultivations were carried out in collaboration with Peter Ödman (Department of Systems Biology, DTU) in a phosphate limited, minimal medium. In the first two batches, in situ NIR spectroscopic measurements were made with a reflectance probe and a transmission probe with 4 mm path length. These results were used for the setup and estimation of a first principles based model. The results are published in Sin et al. (2008)(Appendix B). After the first two cultivations, it was discovered that the reflectance probe was damaged during sterilization due to a construction error and it was sent for repair. The 4 mm transmission probe quickly became saturated and the results could not be used for evaluation of NIR for on-line monitoring. For the remaining seven cultivations, a transflectance NIR probe with adjustable path length was applied along with the reflectance probe (for further details of the different probe types readers are referred Chapter 3). Five batches were used for generation of semi-synthetic samples (SS1-5) in order to break correlations and two batches were run as standard batches (ST2 and ST2). An overview of the nine batch cultivations is presented in table 5.1.

Table 5.1: Overview of *S. coelicolor* batch cultivations.

Cultivations	Comments
FP1, FP2	Standard *S. coelicolor* batch cultivations, results used for the setup and estimation of a first principles based model of *S. coelicolor* (Appendix B)
SS1, SS2, SS3, SS4, SS5	*S. coelicolor* batch cultivations used for generation of semi-synthetic samples. Results used for evaluation of in situ NIR described in 8
ST1, ST2	Standard *S. coelicolor* batch cultivations. Results used for evaluation of in situ NIR described in Chapter 8

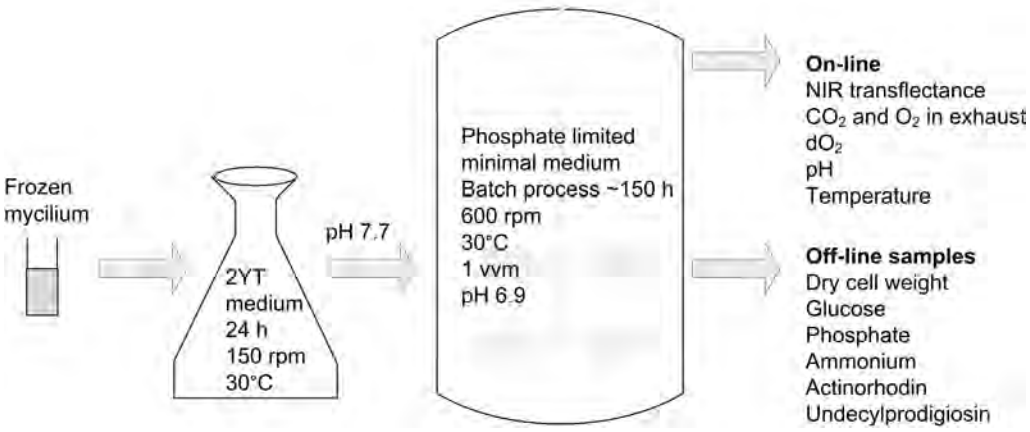


Figure 5.1: Overview of *S. coelicolor* batch cultivation protocol.

The batch cultivations were run after the same protocol (Figure 5.1). Collection of in situ NIR transflectance spectra, exhaust gas analysis and measurement of dissolved oxygen concentration was done on-line. Off-line samples were collected at regular intervals to cover changes in substrate and biomass concentrations. All of the off-line samples were analyzed for dry cell weight, glucose, ammonium, phosphate, actinorhodin and undecylprodigiosin (two antibiotic products). This was done to characterize the growth, substrate consumption and antibiotic production during the cultivation and provide data for model calibration. Furthermore, it was chosen to measure the filtrated samples with off-line NIR transflectance for comparison to the in situ NIR measurements. A detailed description of the medium and analytical techniques is given in Chapter 8 along with a description of the generation of the semi-synthetic samples.

5.3.1 Growth, substrate uptake and antibiotic production in ST1 and ST2

A long lag phase was observed in both ST1 and ST2. The exponential growth phase was almost identical in the two batches with respect to increase in biomass concentration and decrease in glucose and ammonium concentration. Growth ceased around 80 h due to phosphate limitation (data not shown). Approximately 1 L of cultivation broth was lost in ST2 due to foaming. Subsequently, 1 L of fresh medium (without glucose) was added in order to increase the volume to an acceptable level for continued operation. The addition of medium led to a new growth phase around 100 h, which is clearly visible in the biomass, ammonium and dissolved oxygen curves (Figure 5.2). The production of actinorhodin in ST2 is much greater than in ST1, which can be partly explained by the increased biomass concentration.

5.3.2 Correlations in the standard batch and semi-synthetic samples

A significant correlation was observed in the standard batches between the biomass concentration and the glucose concentration, and in particular between the biomass concentration and the ammonium concentration with $r = -0.92$ and $r = -0.97$ respectively (Figure 5.3, top plots). The danger of using a data set with significant correlations for calibration of a NIR probe, is that a model which appears to measure the analyte of interest may in fact be based on signals from the other analytes. In the combined data set of the standard samples and the semi-synthetic samples, the correlations were reduced to $r = -0.76$ between biomass and glucose and $r = -0.79$ between biomass and ammonium (Figure 5.3, bottom plots). This thus improved the usability of the data for calibration of the NIR sensor.

5.3.3 NIR spectra

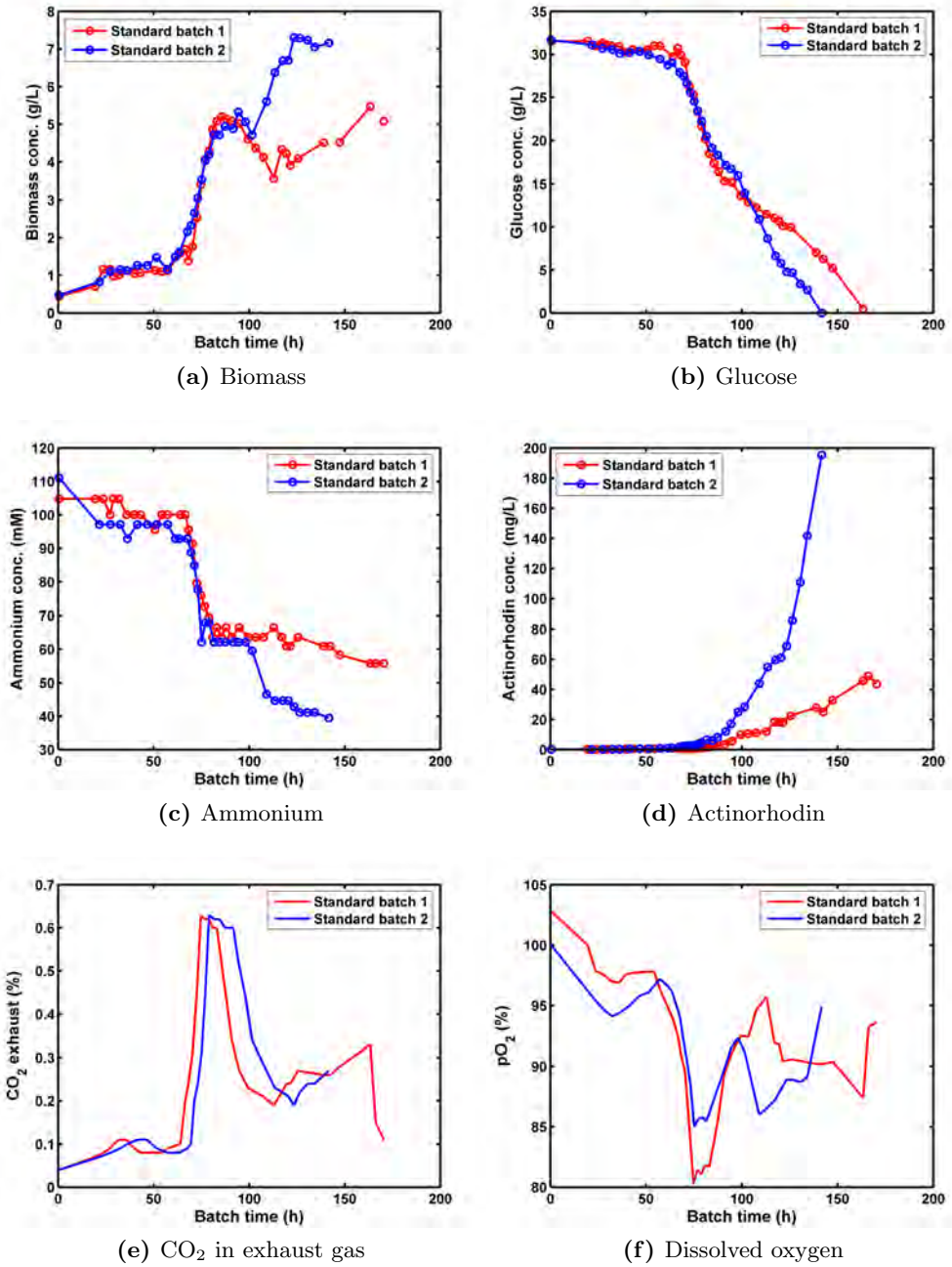
The baseline in the NIR spectra from the in situ transflectance probe varied between 0 and 1.2 absorbance units. Due to the low biomass concentration range, the base line absorbance in the reflectance spectra varied between 1.9 and 2.6.

5.3.4 Evaluation of experiments for NIR calibration

In combination, the standard batch samples and the semi-synthetic samples constitute a data set which is suitable for evaluation of in situ NIR spectroscopy. Due to the low biomass range (0-8 g/L), the signal from the NIR reflectance probe was very low and further analysis was therefore based on the spectra obtained using the transflectance probe. The results are described in Chapter 8.

5.4 *Streptomyces coelicolor* fed-batch cultivations

The *S. coelicolor* fed-batch cultivations were used for evaluation of various online biomass sensors. The main aim of the experimental work was to generate a data set which covered a wide range of operating conditions as well as variation

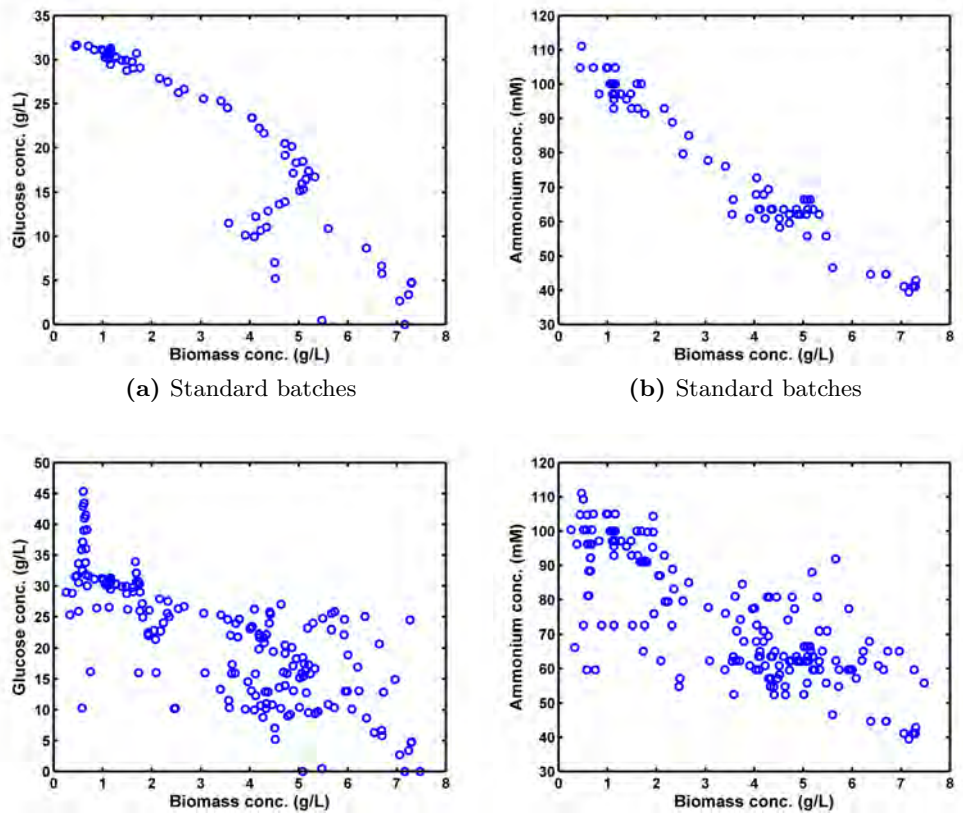


(a) Biomass **(b)** Glucose

(c) Ammonium **(d)** Actinorhodin

(e) CO_2 in exhaust gas **(f)** Dissolved oxygen

Figure 5.2: Development in (a) biomass (g/L), (b) glucose (g/L), (c) ammonium (mM) and (d) actinorhodin (mg/L) concentration in batches ST1 (red) and ST2 (blue). The two bottom plots show the CO_2 level in the exhaust gas and the dissolved oxygen concentration.



(a) Standard batches **(b)** Standard batches

(c) Combined data set **(d)** Combined data set

Figure 5.3: Correlation between biomass and glucose concentrations (left) and between biomass and ammonium concentrations (right). The two top plots show the correlations observed in the two standard batches and the two bottom plots show the corresponding correlations in the combined data set.

in the metabolism and the morphology. Furthermore, it was important to use a procedure which is feasible in an industrial setting. The experiments were therefore designed to imitate an industrial process development and thereby introduce variation by making changes in the cultivation protocol. Process development is an obvious place to implement on-line monitoring. Firstly, the on-line sensors may provide important information regarding the dynamics of the process which is valuable for optimization of the process. Secondly, the process development introduces variation in the data which make it suitable for calibration of sensors. Ideally, on-line sensors could be calibrated during process development and transfered to the production along with the optimized process.

Eight fed-batch cultivations were run as part of process development for improvement of the antibiotic titer. Three factors were changed: feeding mode, pH and medium composition. All cultivations were monitored on-line using dielectric spectroscopy, multi-wavelength fluorescence, NIR spectroscopy, on-line turbidity measurements and standard gas analysis (Figure 5.4). Furthermore, off-line samples were collected at regular intervals and analyzed for dry cell weight, glucose, amino acids, ammonium, phosphate, undecylprodigiosin and actinorhodin in order to characterize the growth and metabolism during the cultivation (Figure 5.4). The resulting data set was used for evaluation of on-line biomass sensors, which is described in Chapter 9. Approximately half of the samples were analyzed using laser diffraction for determination of the particle size distribution and microscopic images were collected for comparison. These data formed the basis of the comparison of laser diffraction and image analysis presented in Chapter 6.

The fed-batch process was designed based on the batch process. A source of amino acids (casamino acids) was added to the medium in order to reduce the lag phase and facilitate the use for spores for inoculation (Figure 5.4). The concentrations of salts, glucose, vitamins were increased to support increased growth. The feed was started after observation of a sustained drop in the carbon dioxide evolution rate. The feed rate was adjusted in order to control the dissolved oxygen concentration according to a predefined profile: Ramp from 80% to 20% over the first 48 h of the fed-batch phase after which the target value remained at 20%. The initial and final volumes of the cultivation broth were 5 L and 9 L respectively. The cultivations were run two at a time. After each round, the results were evaluated and decisions were made for the next cultivations. The initial idea was to do a full factorial design of the three factors at two levels, but observations made during initial cultivations resulted in adjustments of the experimental design. Figure 5.5 summarizes the process development procedure and the decisions taken. Further details of the medium composition and analytical techniques can be found in Chapter 9.

5.4.1 Growth, substrate uptake and antibiotic production

The growth curves are very similar in batches A1-A4 with an almost linear growth from 40 h and onwards (Figure 5.6). Batches B1-2 and C1-2 show more variation in the growth with a fast growth up to 50-60 h (70 h for C1) followed by a period of constant or slowly increasing biomass concentration (Figure 5.6). In batches A1-4 growth stopped around 24 h (observed as a drop in the CER), due to depletion of NH_4^+, after which feed was started (Figure 5.6). A higher level

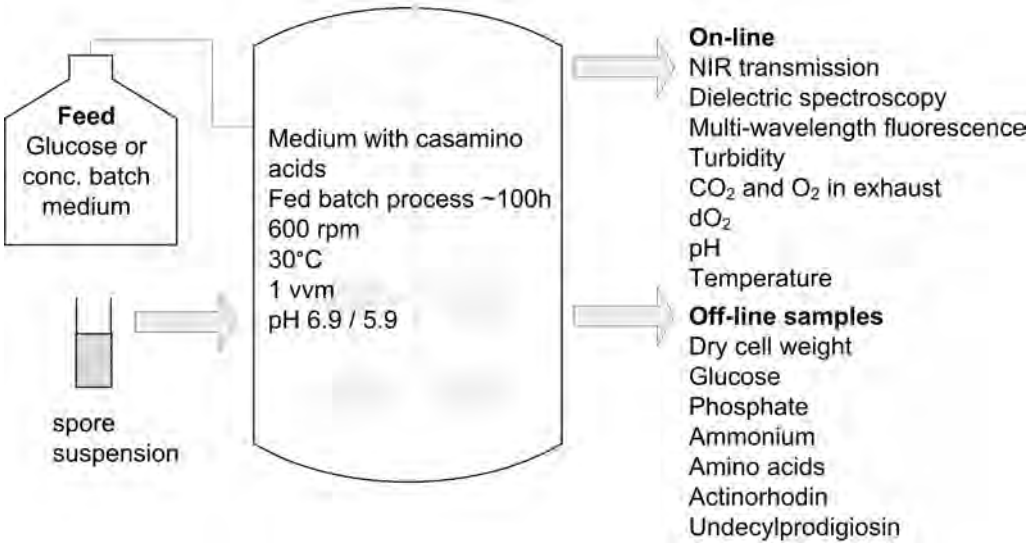


Figure 5.4: Overview of *S. coelicolor* fed-batch cultivations

of NH_4^+ was present in batches B1-2 and C1-2 and growth stopped around 30 h due to depletion of glucose, after which glucose feeding was started. Glucose was limiting in batches A1-3 from around 80 h and in C2 in most of the fed-batch phase. In B1-2 accumulation of glucose was observed after approximately 50 h and in C1 after 60 h. Phosphate was depleted around 80 h in A2-4, around 40 h in B1-2 and around 60 h in C1. Interestingly, growth continued for a period after phosphate depletion. This was also observed in the batch cultivations and may be explained by storage and subsequent utilization of polyphosphates (Sin et al. (2008), Appendix B). Amino acids were available throughout all cultivations (data not shown). The production of actinorhodin was very similar in batches A1-4. Batches B1-2 and C1-2 showed a large variation in the actinorhodin production with almost no production of antibiotics in B2 and C2 and a very large production in C1 (Figure 5.7).

The dissolved oxygen concentration decreased sharply in the batch phase during exponential growth (Figure 5.8). In batch A1-4 the dissolved oxygen concentration continued to decrease in the beginning of the fed-batch phase (up to 45 h for A4 and 55 h for A1-3). In batch C2 the dissolved oxygen concentration decreased down to <10 % during the batch phase. The level remained very low throughout the cultivation due to the high viscosity of the cultivation broth related to the different morphology observed in this batch. In B1-2 and C1 the level stayed above 40 % throughout most of the cultivation. The step wise increase of aeration rate is visible in both the dissolved oxygen concentration and the CO_2 concentration in the exhaust. In batches A1-4 and B1-2 one impeller was placed above the liquid surface in the fermenter. During the fed-batch phase, the liquid level increased above the top impeller causing some splashing. A small increase in the dissolved oxygen concentration was observed when the top impeller was completely covered by the cultivation broth. Finally, a back pressure of 0.5

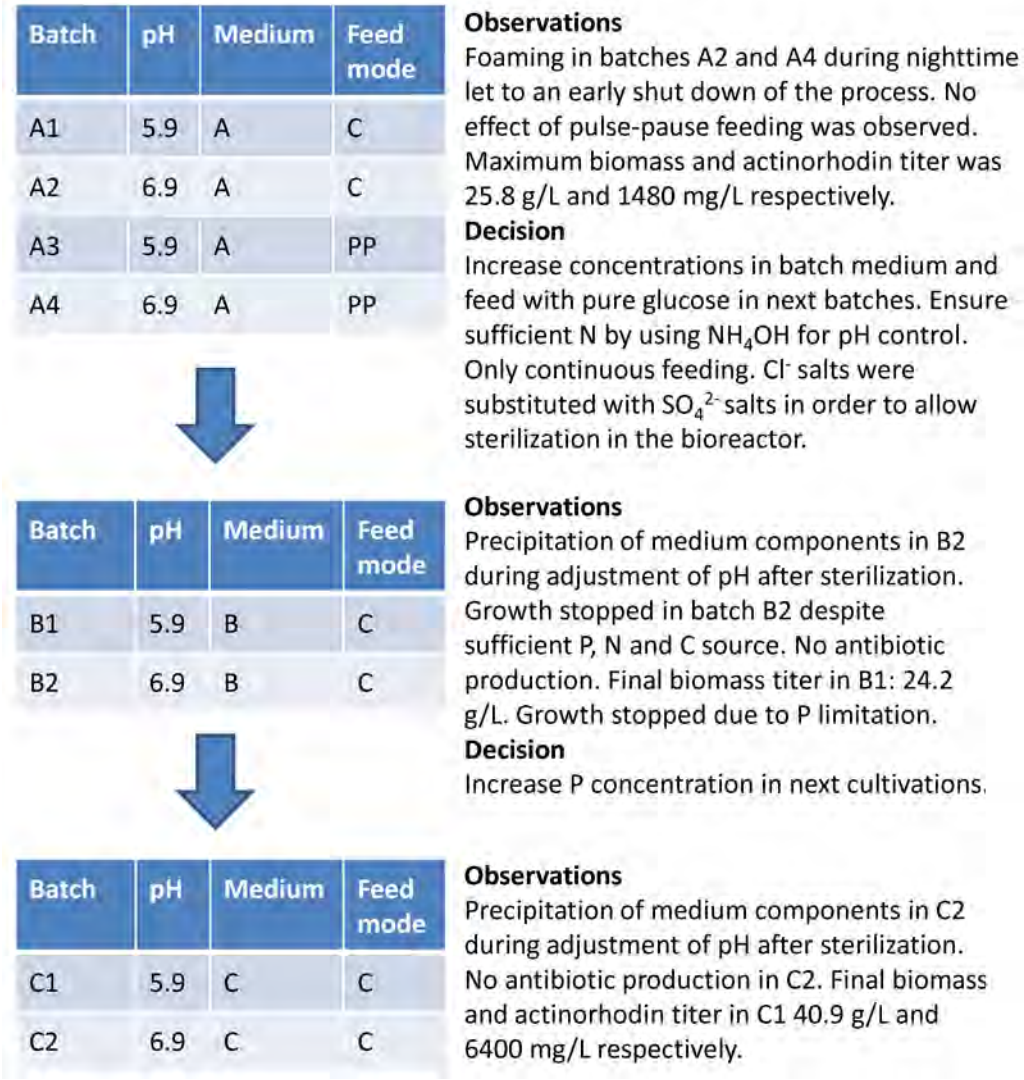
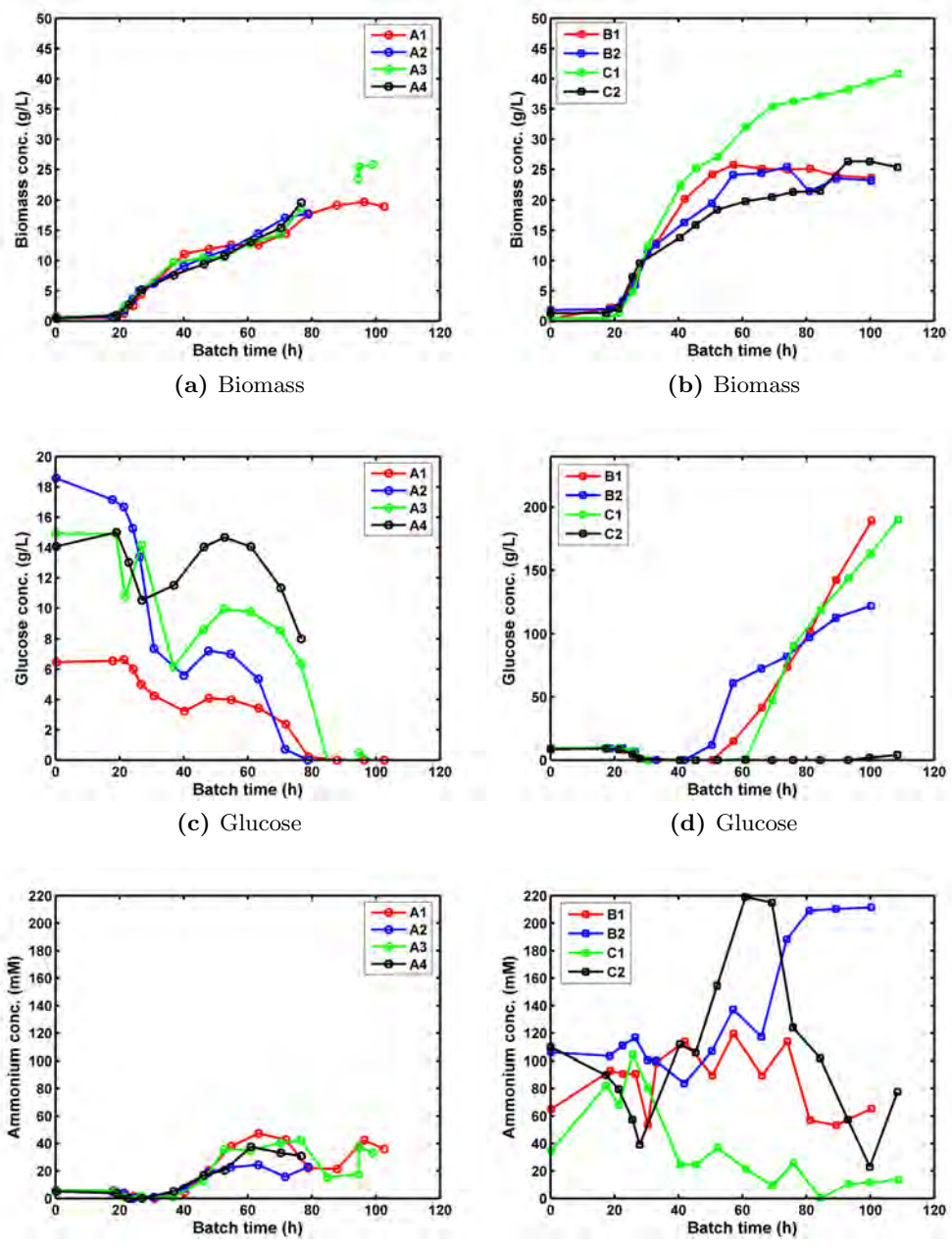


Figure 5.5: Overview of process development for *S. coelicolor* fed-batch cultivations. The exact composition of medium A, B and C is described in Chapter 9. C: Continuous; PP: pulse-pause.

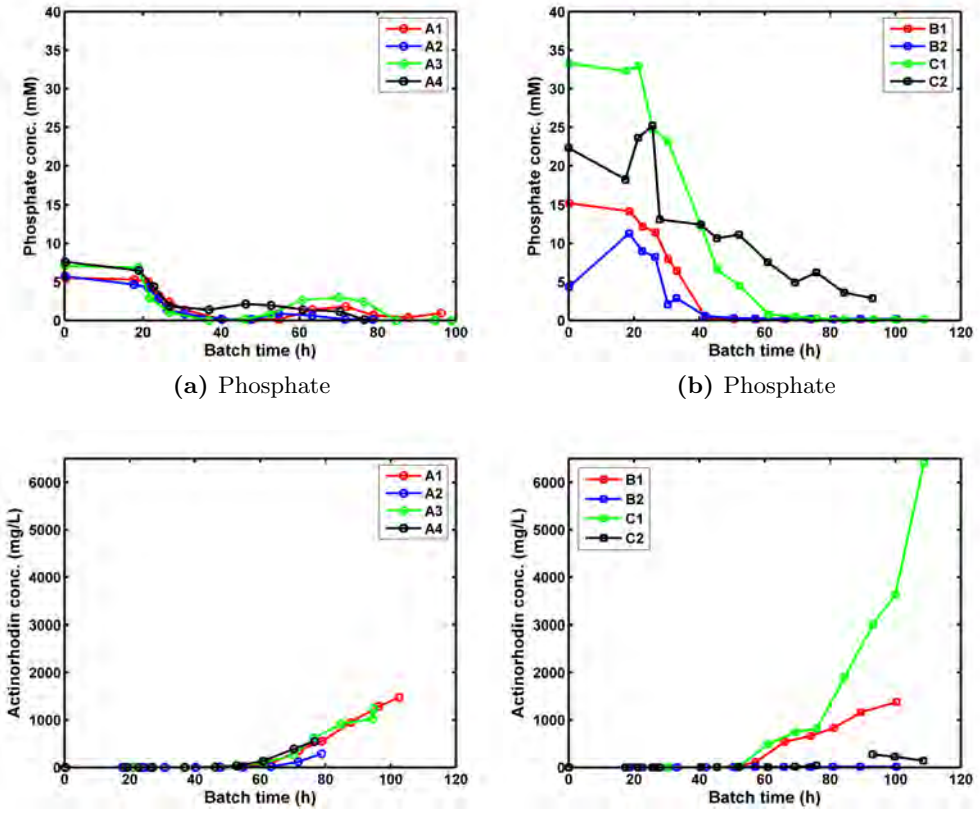


(a) Biomass **(b)** Biomass

(c) Glucose **(d)** Glucose

(e) Ammonium **(f)** Ammonium

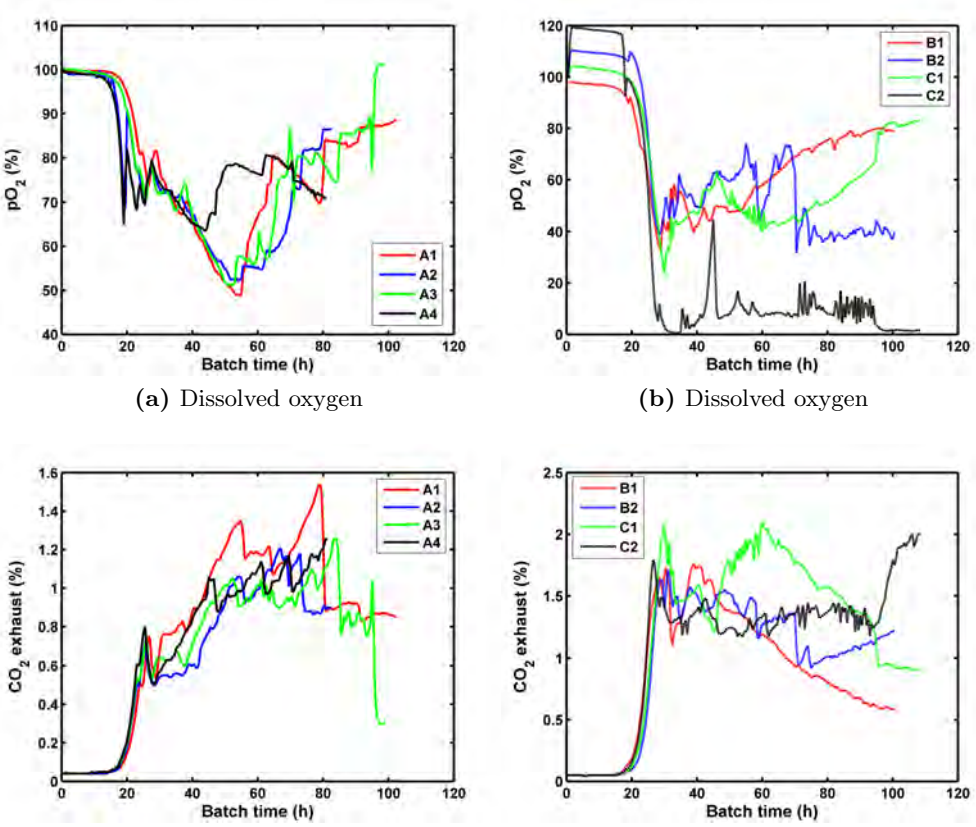
Figure 5.6: Development in biomass (g/L), glucose (g/L) and ammonium (mM) in batches A1-A4 (left) and B1-C2 (right). Note the different axes in (c) and (d).



(a) Phosphate **(b)** Phosphate

(c) Actinorhodin **(d)** Actinorhodin

Figure 5.7: Development in phosphate (mM) and actinorhodin (mg/L) concentration in batches A1-A4 (left) and B1-C2 (right).



(a) Dissolved oxygen **(b)** Dissolved oxygen

(c) Carbon dioxide **(d)** Carbon dioxide

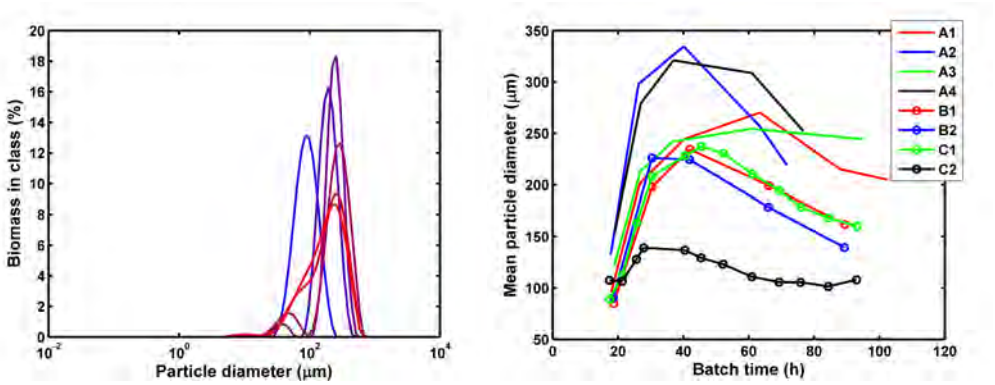
Figure 5.8: Development in dissolved oxygen concentration and carbon dioxide in the exhaust gas in batches A1-A4 (left) and B1-C2 (right).

bar was adjusted manually, but varied throughout and between the cultivations. These effects complicate the interpretation of the dissolved oxygen concentration curves.

The particle size distribution showed a similar development in all batches. In the beginning the cell pellets showed a uniform distribution with an average diameter around 100 μm (Figure 5.9). The pellet diameter increased during exponential growth up to around 220-300 μm at 40 h (except for batch C2 in which the pellet diameter was significantly smaller). At around 40-50 hours a population of small clumps or pellets appeared, which then increased in size and diameter for the remaining part of the cultivation (Figure 5.9). This may be the result of fragmentation of the hyphae, which has previously been coupled to the cell physiology (Stocks and Thomas, 2001). In these experiments the onset of the fragmentation may be coupled to limitation of glucose, phosphate or ammonium.

5.4.2 Evaluation of process development

The feed was adjusted in order to follow a predefined profile for the dissolved oxygen concentration. This worked well during the growth phase. However, in



(a) Size distribution **(b)** Mean particle diameter

Figure 5.9: Particle size distribution measured for samples 2, 5, 7, 10, 13, and 15 in batch A1 (a) and average volume weighted diameter for all fed-batch cultivations (b). The particle size distribution is colored according to the sample number ranging from blue for sample 1 to red for sample 15.

most batches the dissolved oxygen concentration was well above the predefined limit (20 % saturation) in the production phase. The feed rate was therefore constantly at the maximum limit, which let to glucose accumulation in B1-2 and C1. It was concluded, that control of the feed rate based on dissolved oxygen was not suitable for the production phase. The problem of glucose accumulation should of course be addressed in further process development, for example by feeding at a constant rate or perhaps by adjustment of the feeding rate based on glucose measurements using NIR spectroscopic measurements.

It is known that the pH has a large effect on the growth and antibiotic production in *S. coelicolor* (Bystrykh et al., 1996; Elibol, 2002; Kim et al., 2007). Two different pH values (5.9 and 6.9) were tested during process development of the fed-batch process. A pH of 6.8-7 is normally used in house for *S. coelicolor* cultivations. Actinorhodin is poorly soluble at pH lower than 6.8 (Hobbs et al., 1989) and a decrease in production of actinorhodin has been observed at higher pH values (data not shown). A number of studies of *S. coelicolor* cultivations have been carried out a pH values in the range 6.5-7.2, however, many of them do not apply pH control which makes it difficult to evaluate the effect of the pH (Abbas and Edwards, 1990; Coisne et al., 1999; Doull and Vining, 1990; Hobbs et al., 1989). In this study, the highest antibiotic titer and biomass concentration was obtained in batch C1 run at pH 5.9. Due to practical problems in the batches run at pH 6.9 (foaming and medium precipitation), it is not possible to make a fair comparison of the antibiotic production obtained at pH 5.9 and 6.9. However, it can be concluded that the low solubility of actinorhodin at pH 5.9 did not present a problem for the antibiotic production. The cell pellets observed in A2 and A4 run at pH 6.9 were larger (maximum average diameter of 330 μm) compared to the cell pellets observed in A1 and A3 run at pH 5.9 (maximum average diameter 270 μm). However, the low number of batches does not justify any general conclusions.

The change in medium composition A → B → C resulted in an increased growth during the batch phase. In batches B2 and C2, almost no antibiotic production was observed. Furthermore, a highly different morphology was observed in C2 with much lower clump/pellet diameters. These effects are most likely connected to the medium precipitation. Therefore the evaluation of the medium focuses on B1 and C1. Overall, the biomass concentration and antibiotic titers were not improved in batch B1 compared to A1-4. By increasing the phosphate concentration in C1, both the biomass concentration and antibiotic titer were increased. The high antibiotic production in batch C1 can be partly explained by the higher biomass concentration. Furthermore, the combination of limiting phosphate concentration, a relatively low ammonium concentration (<30 mM) combined with excess glucose may provide particularly good conditions for the antibiotic production. In comparison, the ammonium concentration in batch B2 was 70-100 mM. It is known that the availability of P, N and C source has a large influence on the antibiotic production in *S. coelicolor* (Doull and Vining, 1990). Regarding cell morphology, the pellet size in B1 and C1 were generally smaller compared to batches A1-4.

No effect of pulse-pause feeding was observed. The explanation for this is most likely that the cultivations were only nutrient limited in a relatively short period towards the end of the cultivations, i.e. the feed pauses were not reflected in the substrate concentration.

Overall, the process development resulted in an increase in actinorhodin titer from 1.5 mg/L to 6 mg/L which corresponds to 133 mg/g DCW. In comparison, other studies have obtained final titers in the range 10-200 mg/L corresponding to a yield of actinorhodin on biomass in the range 10-53 mg/g DCW (Ates and F., 1997; Ozerginulgen and Mavituna, 1993). Thus the results obtained in the process development described here are highly satisfactory. However, the fed-batch protocol should be further validated and improved.

5.4.3 On-line sensor data

A principal component analysis of the NIR spectra showed that the major variations in the spectra (with and without pretreatment) were not correlated to the biomass concentration as opposed to the measurements with the turbidity probe (at 810-900 nm). Furthermore, clogging of the probe resulting in a dissimilar shape of the spectra was observed in some of the later batches. To avoid this, the measurement window was cleared by adjusting the path length from 1 mm → 3 mm → 1 mm prior to the collection of each sample and corresponding NIR spectrum. This was done in batches B1-C2. Initial analysis of the NIR spectra did not result in successful predictions of biomass concentration. No problems were observed in the collection of data from the gas analyzer, the turbidity probe, the multi-wavelength fluorescence spectroscopic sensor and the dielectric spectroscopic sensor.

5.4.4 Evaluation of experiments for calibration of biomass sensors

The process development resulted in eight batches with great variation in morphology, growth and antibiotic production, i.e. the average diameter ranges from 80 to 340 μm, final biomass concentration ranges from 18 to 41 g/L DCW, and the final titer of actinorhodin ranges from 225 mg/L to 6405 mg/L. These results are representative of an early process development. The collected data set constitute a very challenging and realistic data set for on-line sensor calibration in industry. The results of the biomass sensor calibration are presented in Chapter 9. The data from the NIR sensor were not included in the final comparison, because the observed clogging of the probe would not allow a fair evaluation of NIR spectroscopy for monitoring of biomass concentration.

Bibliography

Abbas AS, Edwards C. 1990. Effects of metals on *Streptomyces coelicolor* growth and actinorhodin production. Applied and Environmental Microbiology 56:675–80.

Amanullah A, Christensen LH, Hansen K, Nienow AW, Thomas CR. 2002. Dependence of morphology on agitation intensity in fed-batch cultures of *Aspergillus oryzae* and its implications for recombinant protein production. Biotechnology and Bioengineering 77:815–826.

Ates M Sand Elibol, F M. 1997. Production of actinorhodin by *Streptomyces coelicolor* in batch and fed-batch cultures. Process Biochemistry 32:273–278.

Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147.

Bhargava S, Wenger KS, Rane K, Rising V, Marten MR. 2005. Effect of cycle time on fungal morphology, broth rheology, and recombinant enzyme productivity during pulsed addition of limiting carbon source. Biotechnology and Bioengineering 89:524–529.

Bibb M. 1996. The regulation of antibiotic production in *Streptomyces coelicolor* A3(2). Microbiology 142:1335–1344.

Bystrykh LV, Fernández-Moreno MA, Herrema JK, Malpartida F, Hodwood DA, Dijkhuizen L. 1996. Production of actinorhodin-related "blue pigments" by *Streptomyces coelicolor* A3(2). Journal of Bacteriology 178:2238–44.

Carlsen M, Spohr AB, Nielsen J, Villadsen J. 1996. Morphology and physiology of an alpha amylase producing strain of *Aspergillus oryzae* during batch cultivations. Biotechnology and Bioengineering 49:266–276.

Chang PK, Ehrlich KC. 2010. What does genetic diversity of *Aspergillus flavus* tell us about *Aspergillus oryzae*? International Journal of Food Microbiology 138:189–199.

Christensen T, Woeldike H, Boel E, Mortensen SB, Hjortshoej K, Thim L, Hansen MT. 1988. High-level expression of recombinant genes in *Aspergillus oryzae*. Bio-technology 6:1419–1422.

Coisne S, Bechet M, Blondeau R. 1999. Actinorhodin production by *Streptomyces coelicolor* A3(2) in iron-restricted media. Letters in Applied Microbiology 28:199–202.

Doull JL, Vining LC. 1990. Nutritional control of actinorhodin production by *Streptomyces coelicolor* A3(2): suppressive effects of nitrogen and phosphate. Applied Microbiology and Biotechnology 32:449–54.

Elibol M. 2002. Product shifting by controlling medium pH in immobilised *Streptomyces coelicolor* A3(2) culture. Process Biochemistry 37:1381–1386.

Hobbs G, Frazer CM, Gardner DCJ, Cullum JA, Oliver SG. 1989. Dispersed growth of *Streptomyces* in liquid culture. Applied Microbial Biotechnology 31:272–277.

Hopwood DA. 1999. Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico. Microbiology 145:2183–2202.

Huge-Jensen B, Andreasen F, Christensen T, Christensen M, Thim L, Boel E. 1989. *Rhizomucor miehei* triglyceride lipase is processed and secreted from transformed *Aspergillus oryzae*. Lipids 24:781–785.

Kieser T. 2000. Practical *Streptomyces* genetics. John Innes Foundation, 613 p.

Kim YJ, Song JY, Moon MH, Smith CP, Hong SK, Chang YK. 2007. pH shock induces overexpression of regulatory and biosynthetic genes for actinorhodin production in *Streptomyes coelicolor* A3(2). Applied Microbial Biotechnology 76:1119–1130.

Kobayashi T, Abe K, Asai K, Gomi K, Juvvadi PR, Kato M, Kitamoto K, Takeuchi M, Machida M. 2007. Genomics of *Aspergillus oryzae*. Bioscience Biotechnology and Biochemistry 71:646–670.

Machida M, Yamada O, Gomi K. 2008. Genomics of *Aspergillus oryzae*: Learning from the history of koji mold and exploration of its future. DNA Research 15:173–183.

Madigan MT, Martinko JM, Parker J. 2003. Brock Biology of Microorganisms. Upper Saddle River, NJ: Pearson Education Inc., 10th edition, 1019 p.

Manteca A, Alvarez R, Salazar N, Yagüe P, Sanchez J. 2008. Mycelium differentiation and antibiotic production in submerged cultures of *Streptomyces coelicolor*. Applied and Environmental Microbiology 74:3877–3886.

Ozerginulgen K, Mavituna F. 1993. Actinorhodin production by *Streptomyces-coelicolor* A3(2) - kinetic-parameters related to growth, substrate uptake and production. Applied Microbiology and Biotechnology 40:457–462.

Pedersen H, 1999. Protein Production by Industrial Aspergilli. Ph.D. thesis, Department of Biotechnology, Technical University of Denmark, Kgs. Lyngby, Denmark.

Pinto L, Vieira LM, Pons MN, Fonseca MMR, C MJ. 2004. Morphology and viability analysis of *Streptomyces clavuligerus* in industrial cultivation systems. Bioprocess and Biosystems Engineering 26:177–84.

Prosser JI, Tough AJ. 1991. Growth mechanisms and growth-kinetics of filamentous microorganisms. Critical Reviews In Biotechnology 10:253–274.

Sin G, Ödman P, Petersen N, Lantz AE, Gernaey KV. 2008. Matrix notation for efficient development of first-principles models within PAT applications: Integrated modeling of antibiotic production with *Streptomyces coelicolor*. Biotechnology and Bioengineering 101:153–171.

Stocks S, Thomas C. 2001. Viability, strength, and fragmentation of *Saccharopolyspora erythraea* in submerged fermentation. Biotechnology and Bioengineering 75:702–709.

Wright LF, Hopwood DA. 1976. Actinorhodin is a chromosomally-determined antibiotic in *Streptomyces coelicolor* A3(2). Journal of General Microbiology 96:289–97.

Chapter 6

Quantification of filamentous morphology of *Streptomyces coelicolor* by laser diffraction

6.1 Introduction

This chapter presents a short study, which was aimed at evaluating laser diffraction for quantification of the morphology (clumps and pellets) of *Streptomyces coelicolor* by comparison to image analysis. Filamentous organisms are used in the production of a wide variety of products such as organic acids, antibiotics and enzymes (Grimm et al., 2004; van Wezel et al., 2006). The morphology of filamentous organisms is closely linked to the rheological properties of the cultivation broth, thus affecting mixing and transfer of heat and mass (Riley et al., 2000). Furthermore, the morphology has a large influence on the productivity and different morphologies i.e. dispersed mycelia or dense pellets, may be optimal in different processes (Riley et al., 2000). The filamentous morphology has therefore attracted much attention over the past years (Bhargava et al., 2003; Clarke, 1962; Hosobuchi et al., 1993; Lin et al., 2010; Nielsen et al., 1995; Petersen et al., 2008; Smith and Calam, 1980).

The morphology of filamentous organisms is traditionally characterized by analysis of microscopic images (Cox et al., 1998; Thomas, 1992). The technique can provide detailed information concerning cell morphology including variables such as total number of tips and average total hyphal length (Li et al., 2002), total projected area, amount of clumps and freely dispersed mycelium (Bhargava et al., 2003; Cox and Thomas, 1992), frequency distribution of pellet size, mean particle size, shape of the pellet (Reichl et al., 1992), clump area, perimeter, compactness and roughness (Tucker et al., 1992), and classification of pellets into smooth and hairy types (Cox and Thomas, 1992). For detailed information on microscopy the readers are referred to Cox et al. (1998). The collection and analysis of images can be fully automated, but the preparation of microscope slides is most often done manually although some on-line and in situ applications exist (Cox and Thomas, 1992; Pearson et al., 2003; Suhr et al., 1995; Treskatis et al., 1997). The preparation of representative slides, which capture both the large pellets and small fragments of the mycelia, is complicated and an unwanted selectivity is often introduced in this step. Very large pellets increase the risk of overloading the slide and as a result the biomass particles will be unevenly distributed. Moreover, the morphology may be altered when the cover glass is pressed over the sample, which of course will result in inaccurate measurements. Alternatively, a cavity slide can be used to preserve the three dimensional structure. However, this is not straight forward because some of the objects will become out of focus. The morphology of large pellets (diameter > 100 μm) have also been determined by

examining submerged cultures on petri dishes without microscopic magnification (Liao et al., 2007; Liu et al., 2008; Moreira et al., 1996; O'Cleirigh et al., 2003; Zmak et al., 2006).

Recently, an in-line particle size analyzer based on focused beam reflectance measurements has been applied for monitoring morphology of filamentous organisms. Pearson et al. (2003) compared image analysis with focused beam reflective measurements and found that changes in cell concentration, aggregate size, and the nature of the aggregate were all reflected in changes in the measured chord length distribution, given by the in-line instrument. A few studies have employed the focused beam reflectance measurements to follow the aggregation of *Aspergillus niger* conidia (Grimm et al., 2004; Lin et al., 2008).

A number of alternative methods, such as flow cytometry, the Coulter counter or CASY cell counting technology, exist for quantification of cell size of unicellular organisms (Ansorge et al., 2007; Dabros et al., 2009; Shapiro, 2003). Flow cytometry is a highly advanced method based on illumination of cells by one or more lasers and subsequent analysis of the light scatter and the fluorescence signals. Upon calibration, it is possible to estimate the cell size based on the scattering properties of the cells. The analysis of fluorescence signals allows analysis of a large variety of variables such as cell physiology, apoptosis, and DNA content through the use of fluorescent labels (Mittag and Tarnok, 2009). The Coulter counter and CASY are specialized technologies for cell counting and cell sizing which are based on electric current. All of the above mentioned methods are fully automated and relatively fast, which makes it possible to analyze a large number of cells in a short time. To our knowledge the above mentioned methods have not been used for quantification of cell aggregates. Hopfer et al. (2001) used flow cytometry to monitor the transition from yeast to filamentous growth as well as the germination of moulds. However, the study did not include cell clumps or pellets (Hopfer et al., 2001).

Laser diffraction represents an alternative method for fast determination of filamentous morphology. It is commonly used for size determination of particles (0.1-3000 μm) in liquids, suspensions, emulsions and powders in a wide range of applications such as the production of food, chemicals, and pharmaceuticals (den Ouden and van Vliet, 1997) (tomato concentrate), (McGarvey et al., 1997) (ink) (Kelly et al., 2006) (pharmaceutical). The measurement principle is based on the light scattering properties of particles, which are directly related to their size. Smaller particles scatter the light at wider angles and with low intensity while the larger particles scatter the light at narrow angles and with high intensity. In practice, the (diluted) sample is passed through the beam of a monochromatic light source (usually a laser). The light scattered by the particles at various angles is measured by multi-element detectors (ISO13320, 2009). The interpretation of the produced scattering pattern to a particle size distribution depends on two major operations. Firstly, a mathematical model must be created to describe scattering of light by homogeneous particles (ISO13320, 2009). Secondly, the measured scattering pattern is deconvoluted to a particle size distribution (ISO13320, 2009). The most widely used models are the Mie theory and the Frauenhofer approximation. The Mie theory is based on Maxwells electromagnetic field equations and provides a rigorous solution which is valid for all sizes of

spheres (ISO13320, 2009).When using this theory it is assumed that (ISO13320, 2009):

1. The particles are optically homogeneous, isotropic and spherical (although some special or regular shapes can be considered as well).
2. The particle is illuminated by a plane wave of known wavelength.
3. The refractive index, both the real and the imaginary part, of the particles and the medium surrounding them is known.
4. The particles have no surface charges and no surface currents.

The Frauenhofer approximation was used in the earlier laser diffraction instruments and does not require knowledge of the samples optical properties of the samples. The use of the Frauenhofer approximation assumes that (ISO13320, 2009):

1. The particles absorb the light completely – only diffraction at contour of the particle is considered.
2. All particles have a circular cross-section (although other, regular shapes can also be taken into account.
3. The particle is illuminated by a plane wave of known wavelength.
4. Only diffraction in the near-forward direction is considered.
5. The particle diameter is much larger than the wavelength of the light.

The advantage of the Frauenhofer approximation is that it does not require any knowledge of the properties of the material, thus it is often used for particles with unknown or variable optical properties. However, the Frauenhofer approximation is only valid for large particles ($d >> \lambda$), whereas the Mie theory provides a general solution for small and large particles as well as transparent and opaque particles (ISO13320, 2009). The selection of the model should be based on knowledge of the particle size, real refractive index and absorption (imaginary part of refractive index). For most particles larger than 50 μm and a relative refractive index greater than 1.2, the two techniques provide similar results (ISO13320, 2009). For particles between 2 and 50 μm the agreement is strongly dependent of the complex refractive index. For opaque particles there is generally good agreement, whereas there may be some disagreement for transparent particles (ISO13320, 2009). The use of Mie theory requires reliable values for the optical properties. To aid in the selection of the optical model and determination of the optical properties, a comparison can be made between the calculated size distribution from laser diffraction to theoretical values or another reference method (ISO13320, 2009).

All of the above mentioned assumptions often do not hold for the measurements of filamentous biomass particles. Firstly, the biomass particles are not spherical. The single mycelial filaments are narrow filaments of varying length. The filaments can form loose clumps of various forms or dense pellets. The larger

pellets are often approximately spherical, however, with filaments extending from the surface. For irregular particles, the size is expressed in terms of a spherical equivalent diameter (ISO13320, 2009). However, it has been shown that if the asymmetric particles are flow oriented (parallel to the direction of the flow) it may result in an apparent bimodality of the particle size distribution (Kelly and Kazanjian, 2006). Furthermore, the exact refractive index of biomass particles is most often not known. Finally, microbial cells have a negative surface charge.

The advantages of laser diffraction include the ease and speed of the measurement. Furthermore, the biomass is analyzed dispersed in a liquid phase, which constitutes its natural environment (at least for submerged cultures) and thus no changes to the morphology or selectivity are introduced by sample preparation. The measurement is fully automated and it is therefore to a large extent independent of human errors and judgment. Due to the speed and the automation of the measurements, the resulting distribution is based on measurements of a large number of particles, which increases the statistical significance of the results. Thus even though it has some shortcomings, laser diffraction constitutes an attractive alternative for determination of morphology of microbial biomass.

Laser diffraction has been used to monitor size distribution of cell aggregates in a number of different biological systems such as for the measurement of planktonic cellular aggregates in sea and lake waters (Casamitjana et al., 2002; Mikkelsen, 2002; Schleheck et al., 2009), measurement of cell aggregates (filamentous and non-filamentous) in activated sludge (Li, 2005; Li et al., 2008; Masse et al., 2006; Wu and Wheatley, 2010) and finally for quantification of filamentous morphology in cultivation processes (Lin et al., 2010; Ödman et al., 2009; Petersen et al., 2008). However, to our knowledge, the method has not been compared to traditional image analysis for quantification of filamentous organisms and thus it is uncertain how the measurements are affected by uncertainties in the refractive index, the deviation from sphericity and the surface charge of the biomass particles. The aim of this short study was therefore to compare the results of image analysis with the results from the laser diffraction method for the quantification of *S. coelicolor* cell clumps and pellets.

6.2 Materials and methods

6.2.1 Cultivation samples

Cultivation samples were collected at five different time points, 18, 26, 40, 63 and 72 h, during a *S. coelicolor* fed-batch cultivation (described in Chapter 5). The size distribution of cell clumps and pellets was measured with laser diffraction and simple image analysis based on microscopic images.

6.2.2 Laser diffraction measurements

The laser diffraction measurements were made on a Malvern Mastersizer 2000 with a Hydro SM manual small sample dispersion unit (Malvern Instruments Ltd., Worcestershire, UK). A Helium-Neon laser emitting red light was used for detection of forward scattering, side scattering and back scattering. A solid state light source emitting blue light was used for detection of wide angle forward

and back scattering. The apparatus is capable of measuring particles in the range between 0.02 to 2000 μm depending on material properties. The method for the measurement of *S. coelicolor* pellets was adopted from Stuart Stocks at Novozymes and developed in house by Peter Ödman. The calculation method was based on Mie Theory, in order to be able to measure both small and large particles. However, as it turned out, the particles in the broth were relatively large (the majority of the particles were above 50 μm) and thus it is likely that the Frauenhofer approximation would have yielded similar results. Since the exact refractive index of the *S. coelicolor* particles was not known, it was initially set to 1.52, which is the refractive index of cellulose. The absorption, related to the imaginary part of the refractive index, was set to 0.1. In the Mastersizer, the absorption should be chosen in the range between 0 (totally transparent) and 1 (totally opaque). The values were subsequently confirmed by qualitative comparison of the laser diffraction measurements and microscopic images of *S. coelicolor* in a small number of samples taken from *S. coelicolor* batch cultivations (Chapter 5). The cell pellets were dispersed in water with a refractive index of 1.33. The stirring was set to 1000 rpm to avoid destruction of the biomass. The dispersions were made by adding small amounts of cultivation broth to distilled water in the dispersion unit until a laser saturation of approximately 12-15% was obtained. This corresponded to approximately 0.5-5 mL of sample depending on the biomass concentration. Three successive measurements were made for each sample and the mean was calculated. A volume weighted particle size distribution was calculated automatically in the Mastersizer 2000 Software (Malvern Instruments Ltd., Worcestershire, UK) and subsequently exported to MATLAB (Mathworks Inc, Massachusetts, USA) for further analysis.

6.2.3 Image analysis

The microscopic images were collected with an Olympus BF40 microscope (Olympus, Tokyo, Japan) equipped with a CoolSNAP cf camera (Photometrics, Arizona, USA). Three slides were prepared and a minimum of 50 images were collected at 10x magnification. The images were analyzed semi-automatically by a simple image analysis algorithm developed in MATLAB (Mathworks Inc, Massachusetts, USA) capable of measurement of cell pellets and clumps with a diameter above 20 μm. A cut-off value of 20 μm was used because the collected images focused on clumps and cell pellets and thus were not optimal for quantification of free hyphae. Furthermore, quantification of smaller particles would have required a more advanced image analysis algorithm to distinguish the particles from the dust and other particulate matter in the broth. Since the amount of free hyphae did not appear significant, it was chosen not to work further on the quantification of smaller particles ($<$20 μm). The estimated particle size distribution was subsequently converted to a volume weighted distribution to fit the results from the laser diffraction method.

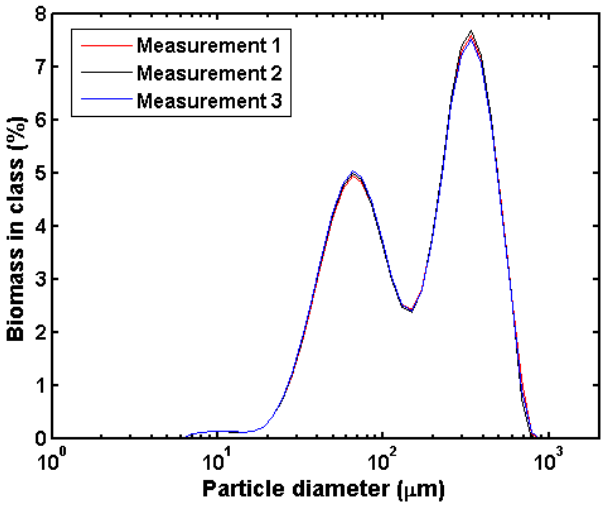


Figure 6.1: Size distributions estimated from three successive laser diffraction measurements of a cultivation sample.

6.3 Results

6.3.1 Repeatability

The three successive laser diffraction measurements taken for each sample resulted in very similar size distributions (Figure 6.1). Based on the three measurements the overall volume weighted mean diameters for the two subpopulations in this particular sample was estimated to 75.9 ± 0.2 μm and 357.1 ± 1.2 μm based on three measurements. Thus, it can be concluded that the laser diffraction method has a very good repeatability.

6.3.2 Comparison to image analysis for quantification of cell clumps and pellets

The cultivation sample taken at 18 h showed a bimodal size distribution with a small (volume weighted) population of particles with a mean diameter of 20 μm and the dominant population with a mean diameter and standard deviation of 137 ± 51.9 μm when analyzed with laser diffraction (Figure 6.2 and Table 6.1). The image analysis identified a population of cell clumps similar to that identified with laser diffraction with a mean diameter and standard deviation of 142 ± 33.3 μm (Figure 6.2a). The smaller particles with approximate diameter of 20 μm were not counted in the image analysis routine due to the set-up, but may be salt crystals or free hyphae (Figure 6.3a).

The laser diffraction measurements of the cultivation samples taken at 26 h and 40 h both showed a unimodal size distribution with mean diameters of 300 μm and 336 μm, respectively (Figure 6.4 and Table 6.1). The corresponding standard deviations were estimated to 85.1 μm and 84.9 μm (Figure 6.4 and Table 6.1). In comparison, the mean diameters and standard deviations estimated by image

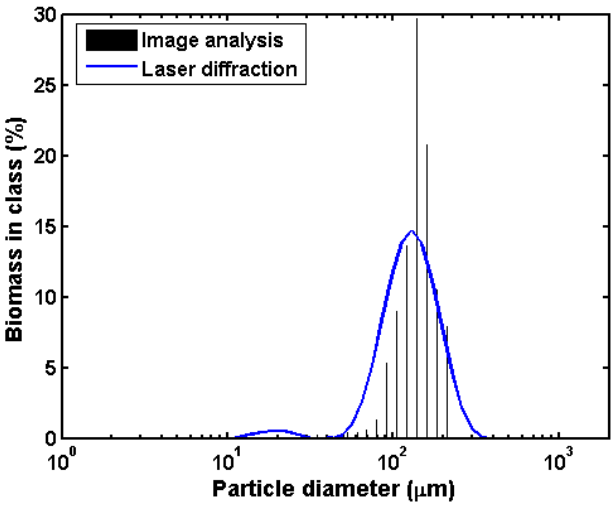
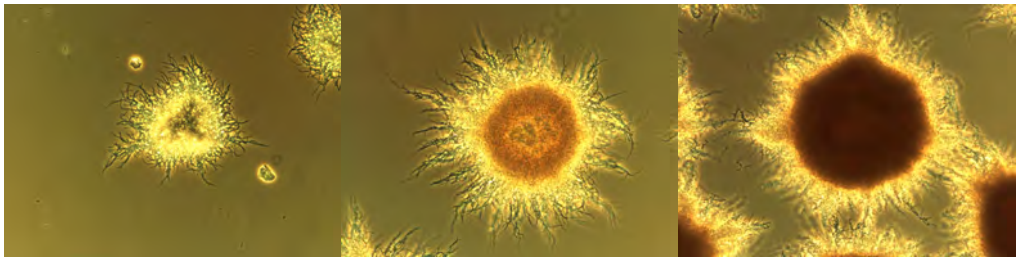
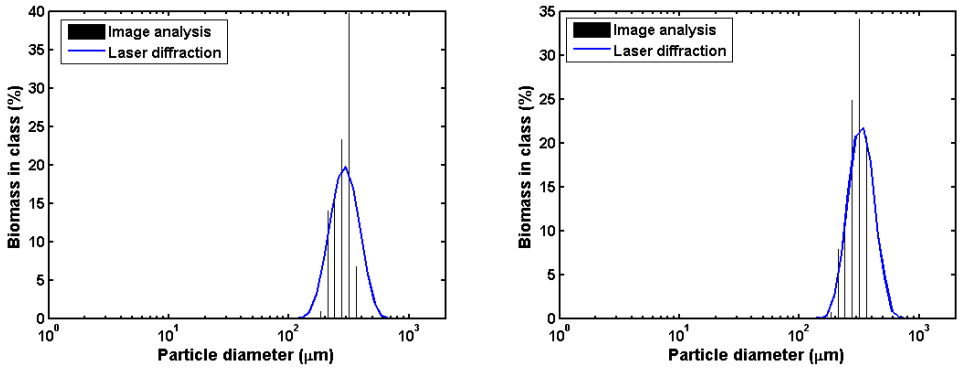


Figure 6.2: Volume weighted particle size distribution measured with laser diffraction (blue line) and image analysis (black bars) for the cultivation sample taken at 18 h.



(a) Sample 18 h **(b)** Sample 26 h **(c)** Sample 40 h

Figure 6.3: Microscopic images collected at 10x magnification in the samples collected at (a) 18 h, (b) 26 h and (c) 40 h.

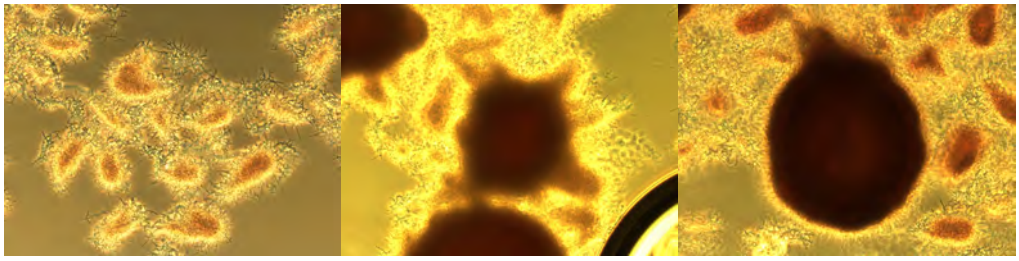


(a) 26 h **(b)** 40 h

Figure 6.4: Volume weighted particle size distribution measured with laser diffraction (blue line) and image analysis (black bars) for the cultivation samples taken at 26 h (a) and 40 h (b).

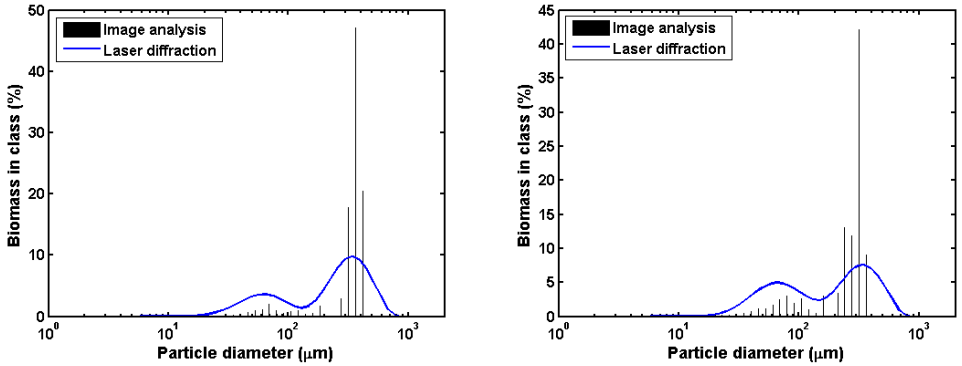
analysis were 282 ± 44.9 and 297 ± 47.2 μm, respectively (Figure 6.3a). Thus, very good agreement was obtained for the mean diameters between image analysis and laser diffraction. However, as it also appears from Figure 6.4, the particle distribution estimated using image analysis appears to be narrower than the distribution resulting from laser diffraction. This is reflected in the relatively large difference between the calculated standard deviations of the two distributions.

For the final two cultivation samples taken at 63 h and 72 h, the laser diffraction measurements showed a bimodal size distribution (Figure 6.5 and 6.6 and Table 6.1). In addition, particles down to a diameter of 6 μm were identified in both samples. These small particles may be free hyphae (Figure 6.5). The size distribution estimated from image analysis also showed a bimodal distribution with mean diameters similar to those identified by laser diffraction (Figure 6.6 and Table 6.1). Again the standard deviation of the populations differed significantly between the results from laser diffraction and image analysis (Table 6.1). Part of the explanation may be that the image analysis results are based on a relatively small number of particles (minimum 50 particles in total) depending on the size. When the particle sizes are transformed to a volume weighted distribution, the few large particles will dominate. Thus the distribution of the sub-population of large particles is based on an even lower number of particles (24 particles with d>120 μm compared to 562 particles with 20 μm<d<120 μm in sample 5). Furthermore, the relative distribution between the two subpopulations varied when image analysis and laser diffraction were applied, i.e. the proportion of large particles (diameter 350 μm) was overestimated using laser diffraction compared to image analysis, whereas the proportion of smaller particles were underestimated compared to image analysis (Figure 6.6). It can be debated which method provides the most realistic estimate of the proportion between small and large particles. The presence of very large pellets resulted in some overloading of the slides for the image analysis. Thus some selection was introduced and the small and large pellets were not evenly distributed. This generally made it difficult to



(a) Sample 63 h (small clumps) **(b)** Sample 63 h (large pellets) **(c)** Sample 72 h

Figure 6.5: Microscopic images collected at 10x magnification in the samples collected at 63 h (a and b), and 72 h (c).



(a) 63 h **(b)** 72 h

Figure 6.6: Volume weighted particle size distribution measured with laser diffraction (blue line) and image analysis (black bars) for the cultivation sample taken at 63 (a) h and 72 h (b).

estimate the true proportion of small and large pellets. Hence, the results should not be taken as evidence that the proportion measured using laser diffraction is either true or false.

Overall, the use of image analysis and laser diffraction for size distribution determination resulted in the identification of similar particle populations and subpopulations. A great similarity was observed between the estimated mean diameters, whereas there was some discrepancy between the standard deviations of the estimated distributions.

6.4 Discussion

One of the largest uncertainties of using laser diffraction for quantification of filamentous morphology is the estimation of the refractive index. The refractive index has a large influence on the size distribution estimated based on laser diffraction (Saveyn et al., 2002). The use of an inappropriate value for the refrac-

Table 6.1: Estimated mean and standard deviation of diameters of subpopulations in the particle size distribution estimated from laser diffraction and image analysis.

	Mean diameter ± std (μm)				
	Sample 18 h	Sample 26 h	Sample 40 h	Sample 63 h	Sample 72 h
Laser diffraction	137±51.9	300±85.1	336±84.9	68.0±34.2 357±142	75.9±41.8 357±143
Image analysis	142±33.3	282±44.9	297±47.2	71.2±26.0 358±48.7	72.3±24.8 292±48.9

tive index can result in an artifact peak of submicron particles in the estimated size distribution as well as inaccurate estimations of the mean particle diameter (Saveyn et al., 2002). However, the acceptable limits of the refractive index are relatively wide. Thus, Saveyn et al. (2002) found no significant differences in weighted mean volume diameter over a very broad range of real-part refractive index values from about 1.55 to 2.10 in the measurement of thiram particles. The refractive index of cellular material of organisms such as *Escherichia coli*, *Staphylococcus albus* and *Staphylococcus aureus*, has previously been estimated to lie in the range between 1.43 and 1.66 (Kononenko et al., 1969). In the same study the refractive index was found to vary slightly with age (±0.04) (Kononenko et al., 1969). These previous findings combined with our results indicate that the chosen value of 1.52, although it may not be completely accurate and it may change with age, is appropriate for measurements of *S. coelicolor* clumps and pellets over the whole range of morphologies in observed in fed-batch operation.

In future applications, the choice of optical model and refractive index should be guided by knowledge of the size and optical properties of the particles. The Frauenhofer approximation, which does not require knowledge of the refractive index, can be applied for large particles. Quantification of smaller particles using the Mie theory requires knowledge of the refractive index of the particles. For quantification of filamentous morphology, values of refractive index will most likely not be available in the literature. Thus the application of laser diffraction will require some study of the optical properties of the biomass particles in order to obtain reliable results. This can for example be done by comparison to a reference method such as image analysis, as described in this study. The samples should preferably cover the range of morphologies expected in future experiments and include variation in other factors, such as age, which are known to affect the optical properties. Alternatively, a suitable refractive index may be chosen by selecting the refractive index that provides the largest estimate of the mean particle diameter as suggested by Hayakawa et al. (1995). The rationale for this is that that choice of an inappropriate of the optical model or of the values of the refractive index may result in significant bias. This bias often manifests itself as inappropriate quantities of material being ascribed to the size classes at the lower end of the size distribution (ISO13320, 2009).

The second major uncertainty of using laser diffraction for quantification of filamentous morphology is the effect of the irregular shaped biomass particles.

Flow orientation of asymmetrical particles with high aspect ratio may result in an apparent bimodality of the particle size distribution, whereas for randomly oriented asymmetrical particles, the apparent volume probability particle size distribution would appear unimodal (Kelly and Kazanjian, 2006; Matsuyama et al., 2000). For the *S. coelicolor* broth, the clumps and pellets are relatively spherical and thus do not constitute a major problem. However, the individual hyphae are asymmetrical with a relatively high aspect ratio and it is likely that the particles will be oriented with the flow to a certain degree. In our samples, the clumps and pellets dominated the volume weighted size distribution and the bimodality observed for some of the samples (>20 μm) was verified by image analysis. We did not observe any problems stemming from the asymmetrical shape of the hyphae. However, for future applications, where the free hyphae constitute a major part of the samples, it would be interesting to perform an additional study on the appearance of hyphae in the size distribution spectra.

In conclusion, image analysis and laser diffraction yielded similar results for the quantification of the *S. coelicolor* cell clumps and pellets with diameters above 20 μm tested in this study. In order to obtain more general results, it would be interesting to make a similar comparison for other organisms. Furthermore, the reliability of the method should always be checked in new applications. Although laser diffraction does not provide the same level of detailed information as image analysis, the method constitutes a fast and reliable alternative to the more laborious image analysis. It is important to mention here, that laser diffraction equipment is already present in many research and industrial facilities, where it is routinely used for other purposes and would therefore not constitute a major investment. Laser diffraction measurements can thus be used to provide information concerning filamentous morphology in studies were detailed image analysis is not possible. Routine measurements of the morphology would help to increase the knowledge of the effects of various process parameters on the morphology. Furthermore, laser diffraction could facilitate studies on the connection between the morphology and productivity. This would be of great importance in the optimization of many filamentous cultivation processes. If a connection exists between the morphology and productivity, control of the morphology would be an important handle to improve productivity.

Bibliography

Ansorge S, Esteban G, Schmid G. 2007. On-line monitoring of infected Sf-9 insect cell cultures by scanning permittivity measurements and comparison with off-line biovolume measurements. Cytotechnology 55:115–24.

Bhargava S, Nandakumar M, Roy A, Wenger K, Marten M. 2003. Pulsed feeding during fed-batch fungal fermentation leads to reduced viscosity without detrimentally affecting protein expression. Biotechnology and Bioengineering 81:341–347.

Casamitjana X, Serra T, Soler M, Colomer J. 2002. A study of the evolution of the particle boundary layer in a reservoir, using laser particle sizing. Water Research 36:4293–4300.

Clarke DS. 1962. Submerged citric acid fermentation of ferrocyanide treated beet molasses: morphology of pellets of *Aspergillus niger*. Canadian Journal of Microbiology 8:133–136.

Cox P, Thomas C. 1992. Classification and measurement of fungal pellets by automated image-analysis. Biotechnology and Biotengineering 39:945–952.

Cox PW, Paul GC, Thomas CR. 1998. Image analysis of the morphology of filamentous micro-organisms. Microbiology 144:817–827.

Dabros M, Dennewald D, Currie DJ, Lee MH, Todd RW, Marison IW, von Stockar U. 2009. Cole-Cole, linear and multivariate modeling of capacitance data for on-line monitoring of biomass. Bioprocess and Biosystems Engineering 32:161–173.

den Ouden F, van Vliet T. 1997. Particle size distribution in tomato concentrate and effects on rheological properties. Journal of Food Science 62:565–567.

Grimm L, Kelly S, Hengstler J, Gobel A, Krull R, Hempel D. 2004. Kinetic studies on the aggregation of *Aspergillus niger* conidia. Biotechnology and Bioengineering 87:213–218.

Hayakawa O, Nakahira K, Tsubaki JI. 1995. Estimation of the optimum refractive index by the laser diffraction and scattering method. Advanced Powder Technology 6:47–61.

Hopfer R, Sparks S, Cox G. 2001. The use of flow cytometry as a tool for monitoring filament formation of fungi. Medical Mycology 39:103–107.

Hosobuchi M, Ogawa K, Yoshikawa H. 1993. Morphology study in production of ML-236B, a precursor of pravastatin sodium, by *Penicillium citrinum*. Journal of Fermentation and Bioengineering 76:470–475.

ISO13320, 2009. Particle size analysis - laser diffraction methods.

Kelly RN, DiSante KJ, Stranzl E, Kazanjian JA, Bowen P, Matsuyama T, Gabas N. 2006. Graphical comparison of image analysis and laser diffraction particle size analysis data obtained from the measurements of nonspherical particle systems. AAPS PharmSciTech 7.

Kelly RN, Kazanjian J. 2006. Commercial reference shape standards use in the study of particle shape effect on laser diffraction particle size analysis. AAPS PharmSciTech 7.

Kononenko AP, Kononenko KI, Mikhov DM. 1969. Dependence of refractive index on physiological state of microbial population. Journal of Applied Spectroscopy 11:795–797.

Li J. 2005. Effects of Fe(III) on floc characteristics of activated sludge. Journal of Chemical Technology and Biotechnology 80:313–319.

Li J, Li Y, Ohandja DG, Yang F, Wong FS, Chua HC. 2008. Impact of filamentous bacteria on properties of activated sludge and membrane-fouling rate in a submerged MBR. Separation and Purification Technology 59:238–243.

Li Z, Shukla V, Wenger K, Fordyce A, Pedersen A, Marten M. 2002. Estimation of hyphal tensile strength in production-scale *Aspergillus oryzae* fungal fermentations. Biotechnology and Bioengineering 77:601–613.

Liao W, Liu Y, Chen S. 2007. Studying pellet formation of a filamentous fungus *Rhizopus oryzae* to enhance organic acid production. Applied Biochemistry and Biotechnology 137:689–701.

Lin PJ, Grimm LH, Wulkow M, Hempel DC, Krull R. 2008. Population balance modeling of the conidial aggregation of *Aspergillus niger*. Biotechnology and Bioengineering 99:341.

Lin PJ, Scholz A, Krull R. 2010. Effect of volumetric power input by aeration and agitation on pellet morphology and product formation of *Aspergillus niger*. Biochemical Engineering Journal 49:213–220.

Liu Y, Liao W, Chen S. 2008. Study of pellet formation of filamentous fungi *Rhizopus oryzae* using a multiple logistic regression model. Biotechnology and Bioengineering 99:117–128.

Masse A, Sperandio M, Cabassud C. 2006. Comparison of sludge characteristics and performance of a submerged membrane bioreactor and an activated sludge process at high solids retention time. Water Research 40:2405–2415.

Matsuyama T, Yamamoto H, Scarlett B. 2000. Transformation of diffraction pattern due to ellipsoids into equivalent diameter distribution for spheres. Particle and Particle Systems Characterization 17:41–46.

McGarvey M, McGregor D, McKay R. 1997. Particle size analysis by laser diffraction in organic pigment technology. Progress in Organic Coatings 31:223–228.

Mikkelsen O. 2002. Examples of spatial and temporal variations of some fine-grained suspended particle characteristics in two danish coastal water bodies. Oceanologica Acta 25:39–49.

Mittag A, Tarnok A. 2009. Basics of standardization and calibration in cytometry–a review. Journal of biophotonics 2:470–81.

Moreira M, Sanroman A, Feijoo G, Lema J. 1996. Control of pellet morphology of filamentous fungi in fluidized bed bioreactors by means of a pulsing flow. Application to *Aspergillus niger* and *Phanerochaete chrysosporium*. Enzyme and Microbial Technology 19:261–266.

Nielsen J, Johansen C, Jacobsen M, Krabben P, Villadsen J. 1995. Pellet formation and fragmentation in submerged cultures of *Penicillium chrysogenum* and its relation to penicillin production. Biotechnology Process 11:93–98.

O'Cleirigh C, Walsh P, O'Shea D. 2003. Morphological quantification of pellets in *Streptomyces hygroscopicus* var. *geldanus* fermentation broths using a flatbed scanner. Biotechnology Letters 25:1677–1683.

Ödman P, Johansen CL, Olsson L, Gernaey KV, Lantz AE. 2009. On-line estimation of biomass, glucose and ethanol in *Saccharomyces cerevisiae* cultivations using in-situ multi-wavelength fluorescence and software sensors. Journal of Biotechnology 144:102–112.

Pearson A, Glennon B, Kieran P. 2003. Comparison of morphological characteristics of *Streptomyces natalensis* by image analysis and focused beam reflectance measurement. Biotechnology Process 19:1342–1347.

Petersen N, Gernaey KV, Stocks S. 2008. Multivariate models for prediction of rheological characteristics of filamentous fermentation broth from the size distribution. Biotechnology and Bioengineering 100:61–71.

Reichl U, King R, Gilles ED. 1992. Characterization of pellet morphology during submerged growth of *Streptomyces tendae* by image-analysis. Biotechnology and Bioengineering 39:164–170.

Riley GL, Tucker KG, Paul GC, Thomas CR. 2000. Effect of biomass concentration and mycelial morphology on fermentation broth rheology. Biotechnology and Bioengineering 68:160–172.

Saveyn H, Mermuys D, Thas O, van der Meeren P. 2002. Determination of the refractive index of water-dispersible granules for use in laser diffraction experiments. Particle and Particle Systems Characterization 19:426–432.

Schleheck D, Barraud N, Klebensberger J, Webb JS, McDougald D, Rice SA, Kjelleberg S. 2009. *Pseudomonas aeruginosa* PAO1 preferentially grows as aggregates in liquid batch cultures and disperses upon starvation. PLoS One 4.

Shapiro HM. 2003. Practical Flow Cytometry. Hoboken, New Jersey: John Wiley and Sons, Inc., 4th edition.

Smith G, Calam CT. 1980. Variations in inocula and their influence on the productivity of antibiotic fermentations. Biotechnology Letters 2:261–266.

Suhr H, Wehnert G, Schneider K, Bittner C, Scholz T, Geissler P, Jahne B, Scheper T. 1995. In-situ microscopy for online characterization of cell-populations in bioreactors, including cell-concentration measurements by depth from focus. Biotechnology and Bioengineering 47:106–116.

Thomas CR. 1992. Image-analysis - putting filamentous microorganisms in the picture. Trends In Biotechnology 10:343–348.

Treskatis S, Orgeldinger V, Wolf H, Gilles E. 1997. Morphological characterization of filamentous microorganisms in submerged cultures by on-line digital image analysis and pattern recognition. Biotechnology and Bioengineering 53:191–201.

Tucker KG, Kelly T, Delgrazia P, Thomas CR. 1992. Fully-automatic measurement of mycelial morphology by image analysis. Biotechnology Progress 8:353–359.

van Wezel GP, Krabben P, Traag BA, Keijser BJF, Kerste R, Vijgenboom E, Heijnen JJ, Kraal B. 2006. Unlocking *Streptomyces* spp. for use as sustainable industrial production platforms by morphological engineering. Applied and Enviromental Microbiology 72:5283–5288.

Wu J, Wheatley A. 2010. Assessing activated sludge morphology by laser and image analysis. Proceedings of the Institution of Civil Engineers-Water Management 163:139–145.

Zmak PM, Podgornik A, Podgornik H, Koloini T. 2006. Impact of pellet size on growth and lignin peroxidase activity of *Phanerochaete chrysosporium*. World Journal of Microbiology and Biotechnology 22:1243–1249.

Chapter 7

Multivariate models for prediction of rheological characteristics of filamentous fermentation broth from the size distribution

The contents of this chapter has been published in Petersen N, Gernaey KV, and Stocks S. 2008. Multivariate models for prediction of rheological characteristics of filamentous fermentation broth from the size distribution. Biotechnology and Bioengineering 100: 61-71.

Abstract

The main purpose of this chapter is to demonstrate that principal component analysis and partial least squares regression can be used to extract information from particle size distribution data and predict rheological properties. Samples from commercially relevant *Aspergillus oryzae* fermentations conducted in 550 litres pilot scale tanks were characterized with respect to particle size distribution, biomass concentration, and rheological properties. The rheological properties were described using the Herschel-Bulkley model. Estimation of all three parameters in the Herschel-Bulkley model (yield stress (τ_y), consistency index (K), and flow behavior index (n)) resulted in a large standard deviation of the parameter estimates. The flow behavior index was not found to be correlated with any of the other measured variables and previous studies have suggested a constant value of the flow behavior index in filamentous fermentations. It was therefore chosen to fix this parameter to the average value thereby decreasing the standard deviation of the estimates of the remaining rheological parameters significantly. Using a partial least squares regression model, a reasonable prediction of apparent viscosity (μ_{app}), yield stress (τ_y), and consistency index (K), could be made from the size distributions, biomass concentration and process information. This provides a predictive method with a high predictive power for the rheology of fermentation broth, and with the advantages over previous models that τ_y and K can be predicted as well as μ_{app}. Validation on an independent test-set yielded a root mean square error of 1.20 Pa for τ_y, 0.209 Pa s^n for K and 0.0288 Pa s for μ_{app}, corresponding to R^2=0.95, R^2=0.94, and R^2=0.95 respectively.

7.1 Introduction

The prediction of suspension rheology is a notoriously difficult and complex specialist area (see Barnes et al. (1989) for an introduction). However, many scientists and engineers today need a practical method for prediction of rheology,

without the need to fully understand theoretical rheology. Examples of this sort of need exist in biotechnology, especially if a scientist or engineer is working the area of mixing and mass transfer in gassed agitated fermentation vessels, a technology widely exploited in the production of a diverse range of products. Fluid viscosity μ appears in many relationships related to mixing and mass transfer, including the Reynolds number and correlations for k_La, the volumetric mass transfer coefficient. Yield stress (τ_y) also appears in correlations used to predict cavern sizes, caverns being the well mixed zones that can form around impellers in otherwise unmixed broth (particularly troublesome in Xanthan gum fermentations (Amanullah et al., 1998)). The use of these useful equations and correlations is presented by Harnby et al. (1992), Tatterson (1991) or Amanullah et al. (2003), for example. Non Newtonian viscosity is usually handled by using μ_{app}, the "apparent viscosity" given by

$$\mu_{app} = \frac{\tau}{\gamma} \tag{7.1}$$

It is most commonly calculated from a simple mathematical approximation of the stress (τ) versus rate (γ) curve at a shear rate determined by what is often referred to as the "Metzner and Otto method".

$$\gamma = K_{MO}N \tag{7.2}$$

In the above equation, K_{MO} is a constant depending on the geometry of the stirrer (-) and N is the impeller speed (rps). The stress (τ) versus rate (γ) curve is normally approximated by the power law,

$$\tau = K\gamma^n \tag{7.3}$$

or the Herschel-Bulkley model,

$$\tau = \tau_y + K\gamma^n \tag{7.4}$$

In the above equation, K is the consistency index (Pa s^n), n is the flow behavior index (-), and τ_y is the yield stress (Pa).

More recently, computational fluid dynamics (CFD) has become popular in the analysis of mixing problems, but without proper rheological information, such models cannot yield relevant results.

Previous work aimed at prediction of viscosity has used traditional empirical power law type correlations to predict some aspect of rheology, eventually resulting in an estimate for μ_{app}; Riley et al. (2000) produced a significant piece of work, reviewing prior art and contributing their own correlation for *Penicillium chrysogenum* (Table 7.1). For the five correlations in Table 7.1, the flow behavior index (n) is fixed or found from other correlations and the consistency index (K) is predicted. In all of the correlations, the biomass concentration features significantly with an index between 1.7 and 2.9, with some other morphological parameter such as maximum dimension, compactness or roughness having less importance with indices in the range -0.96-1.2. Note that the measured morphological parameters are population averages, for a population analyzed by image analysis, where various proportions of the biomass population are ignored based

Table 7.1: Correlations for the prediction of rheology from biomass concentration and population averaged data from image analysis. The table was originally compiled by Riley et al. (2000). Various methods were used to measure rheology. C_m = Biomass concentration (gDCW/L); R = Roughness; C = Compactness; L = maximum dimension.* The original reference uses D to denote the mean maximum dimension.

Source	Correlation	Organism
Tucker and Thomas (1993)	$K = C_m^{2.8} \times R^{0.7} \times C^{1.2} \times const.$	*P.chrysogenum*
Tucker (1994)	$K = C_m^{2.3} \times R^{-0.96} \times C^{0.79} \times 6.6 \times 10^{-5}$	*P.chrysogenum*
Olsvik et al. (1993)	$K = -0.56 + 0.0018 \times R \times C_m^{1.7}$	*A.niger* - continuous
Olsvik and Kristiansen (1994)	$K = 0.38 + 4.8 \times 10^{-5} \times R \times C_m^{2.9}$	*A.niger* - fed-batch
Riley et al. (2000)*	$K = C_m^2 \times 5 \times 10^{-5} \times L - 10^{-3}$	*P.chrysogenum*

on some sensible assumptions, e.g. only clumps are measured since they represent 90% of the total population.

While this traditional log-linear empirical approach to the solving of engineering problems has been at least partially successful, its drawback is that much of the information from the image analysis data is lost by taking a population average, or ignoring parts of the population. For example, each hypha from each part of a size distribution contributes to the viscosity differently, and each hypha of each size class might have a different roughness, contributing to the viscosity slightly differently. Cell debris and other particulate material normally ignored in the image analysis might also be important. Therefore using an arithmetic population mean is perhaps an oversimplification; this is where multivariate modeling might offer advantages.

Recently, multivariate methods, in particular principal component analysis (PCA) and partial least squares regression (PLSR) have become popular in the analysis of spectral data obtained for example by near infrared spectroscopy, and the prediction of material composition. This has been applied in diverse areas, including examples in biotechnology such as quality control of raw materials and on-line composition analysis during submerged fermentation (reviewed for example by Pons et al. (2004) or Scarff et al. (2006)). These statistical methods fall into an empirical category since sets of historical data are needed to fit the model before it can be applied to new spectra, and there is no modeling of the actual physics or chemistry of the system. The analogy between a discretized spectrum, applied to composition analysis, and the possibility of using discretized size distribution data to predict rheology, should be apparent to anyone who is aware of both fields. Multivariate methods have previously been applied to size distribution data in the prediction of rheology in the food industry with respect to milk and orange juice (Carter and Buslig, 1977; Shidara et al., 1995). For an

introduction to PCA and PLSR readers are referred to Chapter 4 in this thesis or Jackson (1980) and Geladi and Kowalski (1986).

Intuitively, one would expect a multivariate approach to the prediction of rheological parameters from particle size distribution data to be successful since one assumes that each particle, whatever its size, will contribute positively to viscosity and so the eventual series of linear relationships in a PLS model has the right basic form. However, rheology is full of caveats; For example it is known that addition of larger spherical particles to a mono-disperse suspension of smaller spheres can actually result in a reduction in viscosity (the Farris effect, (Farris, 1968) - it is therefore not possible to predict a rheological trend with pure intuition. The relationships between particle size, concentration and rheology tend to be increasingly non-linear, and increasingly complex as one starts to consider rigid cylinders rather than spheres (see Barnes et al. (1989), Chapter 7). Add to this complexity, the flexibility of hyphae, the fact that they are normally branched, branch at different frequency depending on the process conditions etc., and one begins to see how a "simple" statistical model might have advantages over a complete theoretical model. Based on successes with milk and orange juice (Carter and Buslig, 1977; Shidara et al., 1995), it was therefore investigated if multivariate techniques could be applied to particle size distribution data and used for prediction of the rheological properties of filamentous fungal broth. The results of doing this are reported in the rest of this article.

7.2 Methods

Data were collected from fermentations of a filamentous fungal strain (*Aspergillus oryzae*) during development of a production process at pilot scale. All fermentations were conducted in fermentation vessels of 550 L total volume. All fermentations ran in fed-batch mode. Precise details cannot be given for commercial reasons, but approximately similar processes have been reported previously (Amanullah et al., 1999, 2002; Bhargava et al., 2003a,b, 2005; Li et al., 2002a,b). The pilot work was focused on fine tuning of the process prior to up-scaling. Different feed strategies were applied which could be described to various degrees as "aggressive" or "conservative" with a feed mode that can be described as "pulse-pause" (constant feed rate interrupted by pauses of approximately 7 minutes) or "continuous" (uninterrupted feeding where the feed rate can be changed continuously). The "pulse-pause" effect is a patent protected method known to reduce the viscosity of the broth through controlled fragmentation or morphological control of the biomass (World patent: WO 2003/029439, Bhargava et al. (2005)).

7.2.1 Analytical methods

Biomass concentration

Biomass concentration was determined in a very traditional way: Pre-weighed 2 cm diameter 0.2 μm filtration membranes were mounted in cartridges; 2 mL of broth in a 5 mL syringe back filled with 3 mL air was passed through and 3-5 mL of deionized water was subsequently used to wash; this procedure was done in duplicate. The filter was dried to constant weight over 24 h at 90^{o}C.

Particle size distribution

Particle size distribution data were gathered with an Explorer/4-Helos (Sympatec) light scattering instrument fitted with a 1000 mm focal length lens. This is a light scattering method and infers a volumetric size distribution, reporting the sizes of spheres with equal volume to the particles actually present. It has the advantage of being much faster than more commonly used microscopy and image analysis techniques (previously used by for example Riley et al. (2000); Thomas et al. (1999); Stocks and Thomas (2001)). Also advantageously, it does not classify particles into "freely dispersed", "clumps" or "pelleted forms" - all particles in the broth are included in the result. This means that it gives a totally unbiased result for size distribution. However, at the same time it fails to report morphological information such as roughness, compactness, or maximum dimension which is obtained by the slower although probably superior technique of image analysis. In the size distribution data the particles are divided into 31 size classes ranging from 4.75 μm to 1610 μm diameter. For each size class the volume percentage of the particles is reported. These data can be transformed by multiplying the volume percentages with the biomass concentration thus obtaining a biomass transformed size distribution.

Rheology

Rheology was characterized with a Carrimed controlled stress rheometer (CSL100). Temperature of the sample was controlled at the fermentation temperature. The rheometer was fitted with a 6 cm, 2^{o} cone truncated by 57 μm. The gap between the cone and the plate was set to 2000 μm in order to cope with the biomass particles in the broth. The software was instructed that a parallel plate was fitted; this is the best analytical approximation to the geometry and gap selected. A parallel plate was not used because it was not otherwise possible to load samples without entrainment of air. This method has been previously reported (Pedersen, 1997), and it is acknowledged that better analytical methods for the characterization of suspension rheology exist (Barnes, 2000) but these were not available at the time of the study. The method, as used, was very reproducible and provided data of sufficient quality for trending and proof of principle presented later in this paper, but it is acknowledged that the results may not be absolute. In all cases a Herschel-Bulkley type fit was made to the data to give a yield stress (τ_y), consistency index (K), and a flow behavior index (n) (Equation 7.4). The parameters were estimated both with and without fixed n. These parameters could then be used to calculate the apparent viscosity (μ_{app}) (Equation 7.1) at any shear rate within the tested range. The apparent viscosity at the shear rate expected in the fermenters was calculated by the commonly used Metzner and Otto method (see for example Cooke et al. (1988); Equation 7.2).

A known problem with power law type models is the large correlation between the parameters, which increases the uncertainty of the parameter estimates. To quantify this, the correlation matrix of the parameter estimates was calculated for a number of samples based on the inverse Fisher Information Matrix as described in Omlin and Reichert (1999) and Seber and Wild (2003). Furthermore a bootstrap analysis (Efron, 1979) was performed to get a quantitative measure

of the uncertainty of the parameter estimates with and without fixed n. In the bootstrap analysis the shear rate/shear stress data sets from the chosen region 0-300 s^{-1} were re-sampled with replacement 200 times using vector re-sampling (Hjorth, 1994). The Herschel-Bulkley model was fitted to each of these re-sampled data sets and based on these values the means and standard deviations of the model parameters were estimated.

7.2.2 Data analysis

PCA and PLS modeling was performed using the PLS toolbox from Eigenvector Research, Inc (USA). PCA was performed of the raw and biomass transformed size distribution data both with and without autoscaling. Individual PLS models were calibrated for prediction of yield stress (τ_y), consistency index (K) and the apparent viscosity (μ_{app}) (Equation 7.4). For the PLS modeling the data set was split into a calibration set (9 batches, 75 samples) and a validation set (3 batches, 24 samples). The batches were chosen such that both continuous and pulse-pause batches were represented in the two data sets. The independent variables used in the PLS models were the fermentation time, feed mode, biomass concentration and each discrete interval of the raw and the transformed size distribution. A value of 1 was arbitrarily chosen to represent pulse-pause feeding and -1 to represent continuous feeding. The best model and the number of components were selected based on a "leave-one-batch-out cross-validation". Finally, the chosen model was tested by test-set validation.

7.3 Results

7.3.1 Data

The total data set consists of 99 samples from 12 fermentations of *Aspergillus oryzae* run under different conditions and with different feeding strategies. For each sample the biomass concentration, the batch time, the feed mode, the size distribution and shear rate/shear stress data were collected. As an example, the size distributions of three different samples from a randomly selected batch are plotted in Figure 7.1.

The full Herschel-Bulkley model provides a good fit of the experimental shear rate/shear stress data with R^2 values between 0.9924-0.9999. As an example the shear rate/shear stress data and the fitted Herschel-Bulkley model are plotted for an arbitrary fermentation sample in Figure 7.2. The correlation matrix of the parameter estimates is shown for the same sample in Table 7.2. It is clear that the parameters are highly correlated with correlation coefficients close to one or minus 1. The highest correlations are found between the flow behavior index n and the consistency index K as expected.

No correlations were found between the estimated flow behavior index and the measured variables: biomass concentration, batch time, feed mode or the size distribution (data not shown). Rather it appeared to be randomly distributed with mean 0.4131 and standard deviation 0.0800. It therefore seemed reasonable to fix n to a constant value of 0.4131 to minimize the uncertainty in the remaining parameters. The fit of the Herschel-Bulkley model with fixed n was still very good

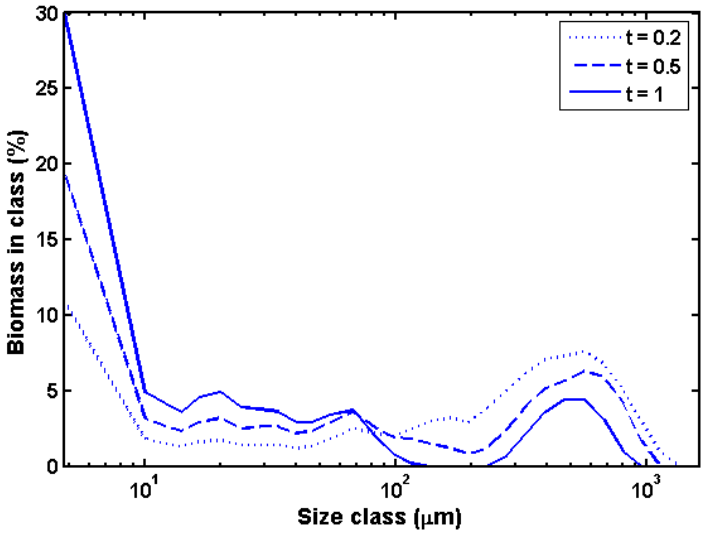


Figure 7.1: Particle size distributions of three samples taken at three different time points (0.2, 0.5, and 1 normalized batch time) during an *A. oryzae* fermentation. The graph shows the volume percentage of the particles in each of 31 size classes ranging from 4.75 to 1,620 mm diameter.

Table 7.2: Correlation matrix of the parameter estimates in the full Herschel-Bulkley model calculated based on the Fisher Information Matrix.

	τ_y	K	n
τ_y	1.0000	-0.9553	0.9268
K	-0.9553	1.0000	-0.9959
n	0.9268	-0.9959	1.0000

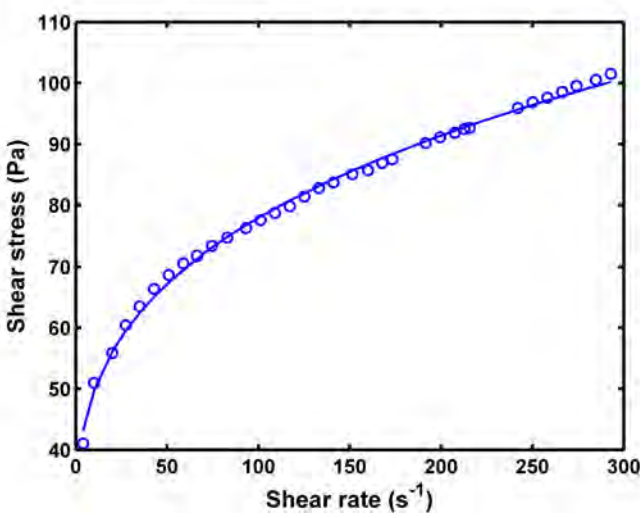


Figure 7.2: Shear rate plotted against shear stress for an arbitrarily chosen fermentation sample. Circles: experimental data; line: full Herschel-Bulkley model approximation.

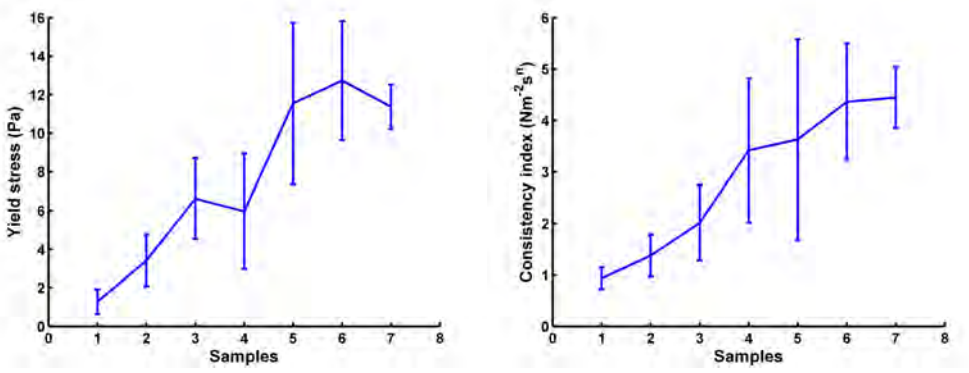
with R^2 between 0.9835-0.9994.

Resulting from the bootstrap analysis, a comparison of the mean and standard deviation of the model parameters estimated with and without a fixed flow behavior index n is shown for an arbitrary batch in Figure 7.3. From Figure 7.3 it is evident that fixing the flow behavior index n has dramatically decreased the variance of the estimated parameters. In addition, it was noticed that the choice of the range of shear rates used for fitting the model had a noticeable effect on the parameter estimates. These findings illustrate a general problem with uncertainty in the estimation of parameters in power law models like the Herschel-Bulkley model. The estimated values of the yield stress and the consistency index are highly sensitive to small changes in the flow behavior index (the exponent) and one should be careful when using the estimated parameters out of the context.

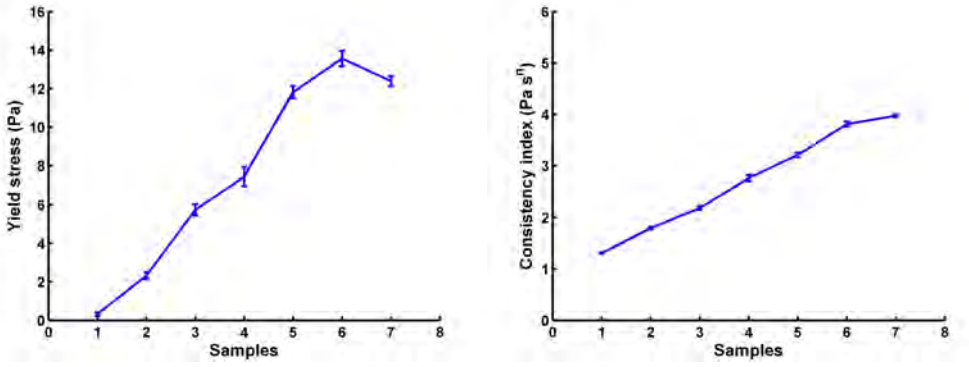
When apparent viscosity was calculated based on the estimates of τ_y, K and n, the bootstrap analysis showed that the certainty of the apparent viscosity is *not affected by the uncertainty in these parameters*. The standard deviation of estimates of μ_{app} was very low both with and without fixed n. A comparison of the apparent viscosity estimated with and without a fixed n showed that the estimated values are almost identical (Root Mean Square Error (RMSE)=0.0047 Pa s). It was therefore decided to use the parameters estimated with fixed n in the further analysis.

7.3.2 Principal Component Analysis

A principal component analysis was performed to explore the size distribution data. The PCA of the raw and biomass transformed size distribution data showed the same overall tendencies; the autoscaling resulted in a larger spread of the observations on the first and second component. The presentation of the results

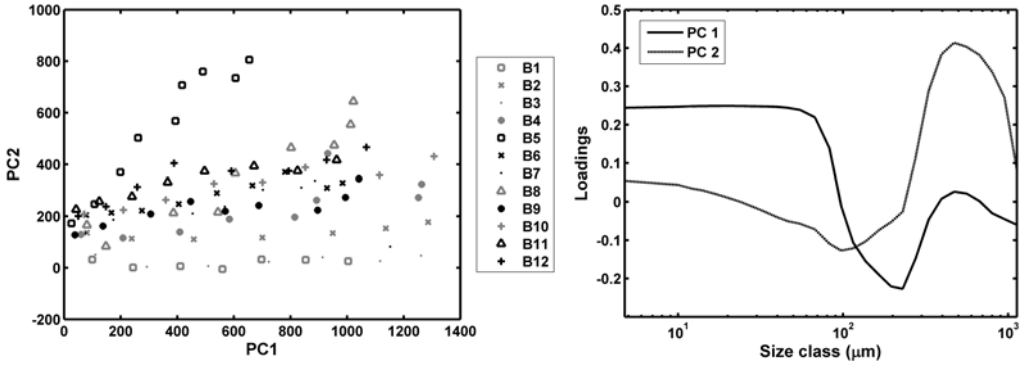


(a) Full model **(b)** Full model



(c) Fixed n **(d)** Fixed n

Figure 7.3: Mean and standard deviation of the estimates of the yield stress (τ_y) and consistency index (K) resulting from bootstrap analysis with vector resampling of one fermentation data set. The parameters were estimated in the full Herschel-Bulkley model (a and b) and in the Herschel-Bulkley model with fixed n (c and d).

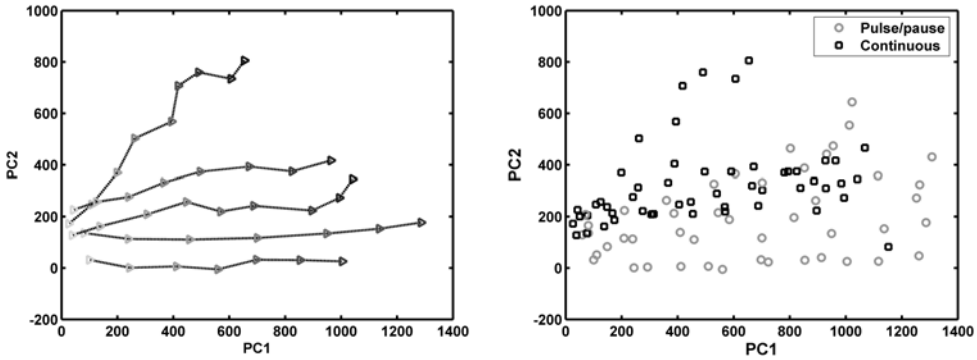


(a) PCA scores **(b)** PCA loadings

Figure 7.4: Scores and loadings for the first two principal components in a PCA model of transformed and auto-scaled size distribution data. (a): Score plot in which the samples are marked according to the batch number. Gray: pulse-pause; black: continuous. (b): Loading plot of the first (-) and second (–) principal component. The first and second principal component explains 54.6 and 18.1% of the variation respectively.

will focus on the biomass-transformed and autoscaled data. The first and second principal component explains 54.6% and 18.1% of the total variation in the size distribution data respectively. The score and loading plots for the first two principal components are shown in figure 7.4. From the score plots it is seen that the samples from individual batches are placed closely together. It is noted that the samples from batch 5, particularly the late samples, are placed far from the other samples in the score plot. Furthermore, the late samples from batch 5 showed high values of the Q-residuals and the Hotelling's T^2 indicating that batch 5 is an outlier. An inspection of the operational data revealed that a technical malfunction had occurred during batch 5 resulting in a highly aggressive feeding profile, which explains why batch 5 differs from the other batches. There is a clear pattern in the loadings on the first and second component. On the first component the smallest size classes are weighted positively. The loadings decrease to negative values for the larger size classes. The loadings on the largest size classes are close to zero on the first principal component. On the second principal component the largest size classes have the highest loadings.

The score plots contain information about the changes in the size distribution depending on the fermentation time as well as the feed mode. This is particularly clear when the scores are colored according to the fermentation time or the feed mode as seen in Figure 7.5. The first component explains the variations in the size distribution as a function of fermentation time, corresponding to an increase in the fraction of small particles and a decrease in the fraction of middle sized particles as the time progresses (Figure 7.5a). Interestingly, the effect of the feed mode on the size distribution is captured in the second principal component (Figure 7.5b). The samples from the continuously fed batches have high scores in the second component whereas the samples from the batches with pulse-pause



(a) PCA scores **(b)** PCA scores

Figure 7.5: Score plots of the first two principal components in a PCA model of transformed and auto-scaled size distribution data. (a): Scores of the samples from five different batches (batches 1, 2, 5, 9, 11) colored according to fermentation time from gray to black with increasing batch time. The lines connect samples from the same batch. The other eight batches show the same time development as is seen in the plot but are left out to ease the interpretation. (b): Samples are colored according to feeding strategy. Black squares: continuous feed; gray circles: pulse-pause feed.

feeding have low scores in the second component. Compared with the loadings this corresponds to a decrease in the fraction of large particles in the pulse-pause fed batches. These observations support the observations of Bhargava et al. (2003b, 2005), who found that the mean projected area was decreased as a consequence of pulse-pause feeding.

7.3.3 Prediction of rheology parameters using PLS

PLS models were calibrated for each of the three rheology parameters τ_y, K and μ_{app} using different combinations of the independent variables: size distribution, biomass-transformed size distribution, biomass concentration, time and feed mode. The independent variables were mean centered and the dependent variables were mean centered and/or log transformed. The number of components in the models was chosen based on evaluation of the Root Mean Square Error of cross-validation (RMSECV). The results showed that the best models were obtained with log transformation and mean centering of the dependent variables (data not shown). Furthermore, it was clear from the initial PLS modeling that batch 5 was an outlier and could not be described by the same PLS model as the other batches. This batch was therefore removed from the data set. The root mean square errors of leave one batch out cross-validation for the PLS models of log transformed rheology parameters are summarized in table 7.3. For comparison the range of τ_y, K and μ_{app} in the calibration data set is 0.138-35.5 Pa, 0.902-6.33 Pa $s^{0.4131}$, and 0.0700-0.8638 Pa s respectively.

The results show that good predictions of all the rheological parameters τ_y,

Table 7.3: Table summarizing the cross-validation errors of the PLS models of the three rheology parameters estimated with n fixed. The calibration data consisted of eight batches. The RMSECV is calculated by leave one batch out cross-validation. Mean centering of the independent variables and log transformation and mean centering of rheology parameters.

	RMSECV (number of components)		
Independent variables	T (Pa)	K (Pa·s^n)	μ_{app} (Pa·s)
biomass	10.1402 (1)	0.8733 (1)	0.1455 (1)
biomass, time, feed mode	8.0947 (3)	0.6857 (1)	0.1085 (1)
size	8.6786 (4)	0.8329 (1)	0.1380 (1)
size, biomass	5.7887 (6)	0.7363 (5)	0.1078 (6)
size, biomass, time	2.8724 (6)	0.5541 (6)	0.0749 (7)
size, biomass, feed mode	3.3706 (6)	0.4499 (8)	0.0702 (6)
size, biomass, time, feed mode	2.9686 (5)	0.4576 (9)	0.0652 (9)
transformed size	9.1660 (4)	0.6313 (6)	0.0902 (6)
transformed size, time	6.3679 (8)	0.5441 (6)	0.0703 (7)
transformed size, feed mode	9.1635 (4)	0.6303 (6)	0.0901 (6)
transformed size, time, feed mode	6.3428 (8)	0.5435 (6)	0.0701 (7)

Table 7.4: The results of test-set validation of the three chosen PLS models. τ_y: size distribution, biomass and time (6 PCs); K: size distribution, biomass and feed mode (8 PCs); μ_{app}: size distribution, biomass, time and feed mode (9 PCs).

	τ_y (Pa)	K (Pa·s^n)	μ_{app} (Pa·s)
Range	0.1096-38.4064	0.9528-6.5366	0.0749-0.9231
RMSEV	1.8447	0.2086	0.0298
R^2	0.8932	0.9384	0.9441

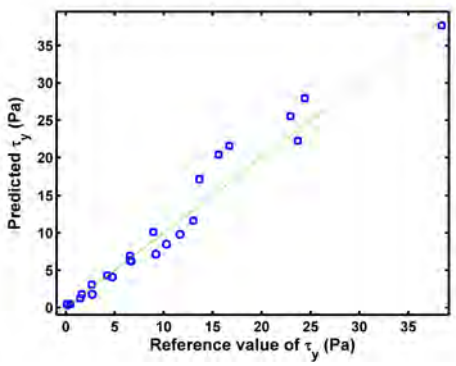
K, and μ_{app} can be obtained with PLS models based on size distribution data, biomass concentration, and the feed mode using between six to eight components. For both μ_{app} and τ_y a small decrease in RMSECV is found by including time as an independent variable. However, as time is not a physical variable and the improvement is very small it is chosen to focus on the models based on size distribution data, biomass concentration, and the feed mode. Compared to the predictions made without size distribution data a clear improvement was achieved by including the size distribution data. This means that a linear model with time or biomass concentration alone is not enough for the prediction of the rheological parameters, and that the size distribution data can be used advantageously. No improvement is found for the models based on the biomass-transformed size distribution data. This indicates that the effects of biomass concentration and the size distribution are additive rather than multiplicative.

The results of the test-set validation are summarized in Table 7.4. The prediction errors are smaller than for the cross-validation. The largest prediction errors are seen for high parameter values where the data are sparse. The predicted values of the parameters are plotted against the reference values in Figure 7.6. The predictions of the yield stress, τ_y, the consistency index, K, and the apparent viscosity, μ_{app}, are all considered highly satisfactory (Figure 7.6).

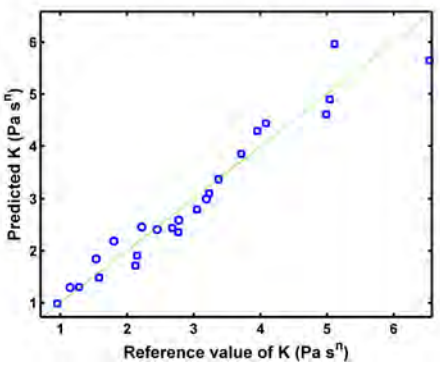
The resulting regression coefficients for biomass concentration and feed mode are found in Table 7.5 while the regression coefficients for the size distribution data are plotted in Figure 7.7. The biomass concentration contributes positively to all of the rheological parameters while a pulse-pause feeding has a decreasing effect on the parameters. Regarding the size classes it is particularly noticeable that the smaller size classes (except the smallest) from 10 to 34 μm seem to have a decreasing effect on the yield stress, consistency index and the apparent viscosity while the size classes from 46 to 115 μm contribute positively to all three rheology parameters. For the larger size classes there does not seem to be a clear pattern.

7.4 Discussion

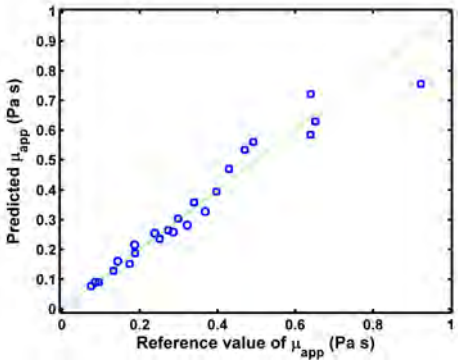
The purpose of this paper was to demonstrate that PCA and PLS can be used to extract information from size distribution data and to predict rheological properties. This approach is a statistical method and is successful in predicting τ_y, K, and μ_{app} from size distribution, biomass concentration data, and feed mode,



(a) The yield stress



(b) The consistency index



(c) The apparent viscosity

Figure 7.6: Predicted values of the rheological parameters resulting from test set validation plotted against the reference values. (a) The yield stress (τ_y); (b) The consistency index (K); (c) The apparent viscosity (μ_{app}). Circles: pulse-pause feeding; squares: continuous feeding.

Table 7.5: Regression coefficients for biomass, time and feed mode model in the chosen PLS models of τ_y, K and μ_{app}.

	Biomass	Time	Feed mode
τ_y	0.0059002	0.0088818	-
K	0.015571	-	-0.056436
μ_{app}	0.010577	0.0035561	-0.037712

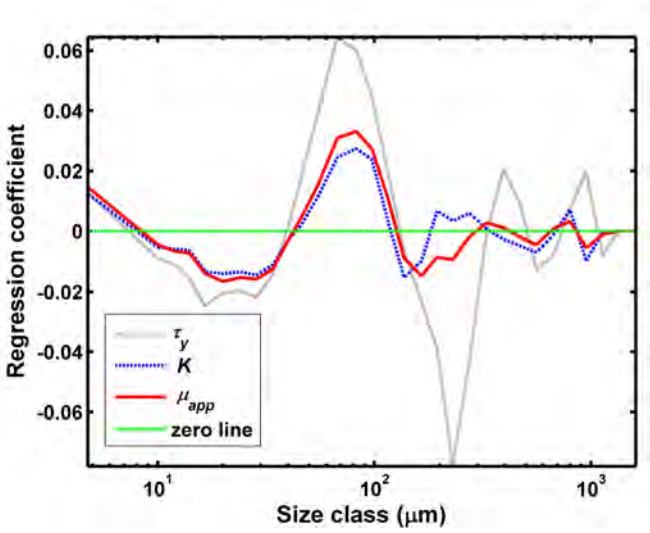


Figure 7.7: Regression coefficients in the chosen PLS models of τ_y, K and μ_{app}.

without modeling the physics of the system. It was shown that with a small increase in the RMSECV of the apparent viscosity, the batch time can be ignored which indicates that the PLS model uses the physical information from the biomass size distribution in a meaningful way. Further, the factor of feed mode (pulse-pause vs. continuous mode), known to influence morphology and viscosity (Bhargava et al., 2003a,b, 2005), improves the predictive power of the model of τ_y, K and μ_{app} indicating that the model is sensitive to the concept of the "morphology-viscosity" type of a strain.

Analysis of the resulting regression model can be used to increase understanding of the correlations between the size of the particles and the rheological parameters. In this work it was found that particles with diameters between 46 and 115 μm contribute positively to both yield stress, consistency index as well as the apparent viscosity. It would be interesting to examine if this is a general tendency by analysis of size distribution data from other scales (lab, production) and other organisms.

The results show that important information is present in the entire range of particles. The multimodal shape of the particle size distribution illustrates that an average size of the particles can be an unsuitable measure of the particle size. For example for the size distributions shown in Figure 7.1 the average diameter would be calculated as 300 μm, 217 μm and 115 μm for the samples taken at the normalized batch times 0.2, 0.5 and 1 respectively. An inspection of the size distributions shows that only a small volume fraction of the particles have a diameter around the calculated averages. In this case the average value of the particle sizes does not really mirror the size of the particles actually present in the broth. This paper therefore encourages the utilization of the entire size distribution spectrum in the analysis of the link between morphology and rheology. The framework is not limited to size distribution data measured using laser light scattering techniques. Advanced automated microscopes are available which are able

to measure a statistically significant size distribution of the particles. In addition to the size distribution, other morphological data will be available from image analysis that could be important for the prediction of rheological parameters.

All batches were run under different conditions and feeding strategies to ensure that the batches span a wide range of operating conditions. However, a technical malfunction occurred in batch 5 and resulted in a feeding strategy that was extremely different from the other batches. It was chosen to exclude this batch in the calibration of the PLS model for the prediction of the rheological parameters. This points to the fact that the calibrated model is only valid for this scale, this particular strain, and within a limited range of operating conditions. This may be caused by the fact that the rheological phenomena are highly non-linear and thus the linear PLS model is an approximation which is only valid within a limited range. Non-linear methods exist which could be tested for prediction of rheological parameters. Furthermore, this sensitivity to strain, scale and operating conditions may also indicate that the presented multivariate model based on size distribution from laser sizing, is only robust for one "morphology-viscosity" type and that there is therefore not enough information in the size distribution data from laser sizing alone to fully predict rheological properties. It is likely, that by combining a superior particle characterization method, such as image analysis, with PLS modeling, the amount of information in the size distribution by including morphological parameters (e.g. compactness, roughness, maximum dimension, hyphal growth unit etc.) would increase the predictive power measurably.

More experimental data and a substantial amount of work is needed to test if the model can be extended to cover different morphologies, scales and organisms. Firstly, this work would provide valuable insight into the effect of morphology on the rheological parameters. Secondly, if successful the size distribution data have potential in a more general model for the prediction of rheology of filamentous fermentation broths. Such a general model would be highly useful in design of fermentation processes. In this case it would also be necessary to be able to predict the size distribution of particles in a fermentation broth. The prediction of fungal morphology is another important area of research which has been considered earlier by Jüsten et al. (1998) among others. In their paper from 1998 they showed that the morphology represented by the mean projected area is correlated with the energy dissipation/circulation function. One could argue that it would be easier to predict the rheological parameters directly rather than predicting the size distribution to predict the rheological parameters. However, it is believed that the size distribution provides a physically meaningful link between the operating conditions, the microorganisms and the rheology.

7.5 Conclusion

It has been shown that chemometric methods can be used to extract valuable information from laser size distribution data. PCA can be used to provide an overview of the multivariate data, explore the data and reveal important correlations. A multivariate PLS model, using laser size distribution data and biomass concentration can be used to model the rheological properties of filamentous

fermentation broth with equal or greater accuracy than traditional engineering power-law type relationships that are based on population average data from image analysis. However, the present model does not extend to different strains or scales. Therefore, while proof of principle for the multivariate approach to modeling rheology has been obtained, better models might combine population data from image analysis, including data on compactness, maximum dimension etc., with the multivariate approach. Furthermore, due to the non-linearity of the rheological phenomena, non-linear methods would possibly be able to cover a wider range of morphologies and scales.

The multivariate approach to prediction of rheology should by no means be confined to the area of filamentous fermentation broth. It could also supply a practical and previously unavailable approach to prediction of rheological properties in many biochemical, pharmaceutical, formulation and engineering situations where the range and diversity of particle morphology and rheological properties in solid-liquid, liquid-liquid, and solid-solid suspensions and mixtures is immense.

Acknowledgement

Acknowledgement is made of Mette Cebolinho Johansen and Fermentation pilot plant staff at Novozymes A/S Denmark for assistance in running the fermentations. Also, to the Enzyme Analytical Lab for performing the size distribution analyses. The PhD project of Nanna Petersen is supported by a grant from the Innovative Bioprocess Technology Research Consortium financed by the Danish Research Council for Technology and Production Sciences, Chr. Hansen A/S, Danisco A/S and Novozymes A/S.

Bibliography

Amanullah A, Blair R, Nienow A, Thomas C. 1999. Effects of agitation intensity on mycelial morphology and protein production in chemostat cultures of recombinant *Aspergillus oryzae*. Biotechnology and Bioengineering 62:434–446.

Amanullah A, Buckland BC, Nienow A. 2003. Mixing in the fermentation and cell culture industries. In: Paul EL, Atiemo-Obeng V, Kresta SM, editors, Handbook of industrial mixing: science and practice, New Jersey: Wiley-Interscience. p 1071–1170.

Amanullah A, Christensen LH, Hansen K, Nienow AW, Thomas CR. 2002. Dependence of morphology on agitation intensity in fed-batch cultures of *Aspergillus oryzae* and its implications for recombinant protein production. Biotechnology and Bioengineering 77:815–826.

Amanullah A, Hjorth S, Nienow A. 1998. New mathematical model to predict cavern diameters in highly shear thinning, power law liquids using axial flow impellers. Chemical Engineering Science 53:455–469.

Barnes HA. 2000. Measuring the viscosity of large-particle (and flocculated) suspensions - a note on the necessary gap size of rotational viscometers. Journal of Non-Newtonian Fluid Mechanics 94:213–217.

Barnes HA, Hutton JF, Walters K. 1989. An introduction to rheology. Elsevier Science Publishers B.V.

Bhargava S, Nandakumar M, Roy A, Wenger K, Marten M. 2003a. Pulsed feeding during fed-batch fungal fermentation leads to reduced viscosity without detrimentally affecting protein expression. Biotechnology and Bioengineering 81:341–347.

Bhargava S, Wenger KS, Marten MR. 2003b. Pulsed addition of limiting-carbon during *Aspergillus oryzae* fermentation leads to improved productivity of a recombinant enzyme. Biotechnology and Bioengineering 82:111–117.

Bhargava S, Wenger KS, Rane K, Rising V, Marten MR. 2005. Effect of cycle time on fungal morphology, broth rheology, and recombinant enzyme productivity during pulsed addition of limiting carbon source. Biotechnology and Bioengineering 89:524–529.

Carter RD, Buslig BS. 1977. Viscosity and particle size distribution in commercial Florida frozen concentrated orange juice. United States of America, Florida State Horticultural Society [Symposium] and Proceedings of the Florida State Horticultural Society .

Cooke M, Middelton JR, Bush JR. 1988. Mixing and mass transfer in filamentous fermentations. In: Proceedings of 2nd International conference on bioreactors. Cambridge, UK: BHRA.

Efron B. 1979. Bootstrap methods: Another look at the Jackknife. Annals of Statistics 7:1–26.

Farris RJ. 1968. Prediction of viscosity of multimodal suspensions from unimodel viscosity data. Transactions of The Society of Rheology 12:281–301.

Geladi P, Kowalski BR. 1986. Partial least squares regression: a tutorial. Analytica Chimica Acta 185:1–17.

Harnby N, Edwards MF, Nienow AN. 1992. Mixing in the process industries. Oxford, UK: Butterworth-Heinemann Ltd.

Hjorth JSU. 1994. Computer intensive statistical methods. London, UK: Chapman and Hall, 280 p.

Jackson JE. 1980. Principal components and factor analysis: Part I - principal components. Journal of Quality Technology 12:201–213.

Jüsten P, Paul G, Nienow A, Thomas C. 1998. Dependence of *Penicillium chrysogenum* growth, morphology, vacuolation, and productivity in fed-batch fermentations on impeller type and agitation intensity. Biotechnology and Bioengineering 59:762.775.

Li Z, Shukla V, Wenger K, Fordyce A, Pedersen A, Marten M. 2002a. Effects of increased impeller power in a production-scale *Aspergillus oryzae* fermentation. Biotechnology Progress 18:437–444.

Li Z, Shukla V, Wenger K, Fordyce A, Pedersen A, Marten M. 2002b. Estimation of hyphal tensile strength in production-scale *Aspergillus oryzae* fungal fermentations. Biotechnology and Bioengineering 77:601–613.

Olsvik E, Kristiansen B. 1994. Rheology of filamentous fermentations. Biotechnology Advances 12:1–39.

Olsvik E, Tucker D, Thomas C, Kristiansen B. 1993. Correlation of *Aspergillus niger* broth rheological properties with biomass concentration and the shape of mycelial aggregates. Biotechnology and Bioengineering 42:1046–1052.

Omlin M, Reichert P. 1999. A comparison of techniques for the estimation of model prediction uncertainty. Ecological Modelling 115:45–59.

Pedersen AG. 1997. k_La characterization of industrial fermentors. In: Nienow AW, editor, 4th International conference on bioreactor and bioprocess fluid dynamics. UK: Mechanical Engineering Publications Limited.

Pons MN, Bonte S, Potier O. 2004. Spectral analysis and fingerprinting for biomedia characterisation. Journal of Biotechnology 113:211–230.

Riley GL, Tucker KG, Paul GC, Thomas CR. 2000. Effect of biomass concentration and mycelial morphology on fermentation broth rheology. Biotechnology and Bioengineering 68:160–172.

Scarff M, Arnold S, Harvey L, McNeil B. 2006. Near infrared spectroscopy for bioprocess monitoring and control: Current status and future trends. Critical Reviews in Biotechnology 26:17–39.

Seber G, Wild C. 2003. Nonlinear regression. New York, USA: John Wiley and Sons, 768 p.

Shidara H, Endo M, Chida T, Watanabe R, Kamiwano M. 1995. Particle size distribution after homogenization and viscosity of fresh cream. Journal of Japanese Society of Food Science and Technology [Nippon Shokuhin Kogyo Gakkaishi] 42:230–236.

Stocks S, Thomas C. 2001. Viability, strength, and fragmentation of *Saccharopolyspora erythraea* in submerged fermentation. Biotechnology and Bioengineering 75:702–709.

Tatterson GB. 1991. Fluid Mixing and gas dispersion in agitated tanks. Library of Congress, 564 p.

Thomas C, Sebastine I, Cox P, Stocks S. 1999. Characterisation of percentage viability of *Streptomyces clavuligerus* using image analysis. Biotechnology Techniques 13:419–423.

Tucker KG, 1994. Relationship between mycelial morphology biomass concentration and broth rheology in submerged fermentations. Ph.D. thesis, University of Birmingham.

Tucker KG, Thomas CR. 1993. Effect of biomass concentration and morphology on the rheological parameters of *Penicillium chrysogenum* fermentation broths. Food and Bioproducts Processing 71:111–117.

Chapter 8

In situ NIR for monitoring of glucose and ammonium in *Streptomyces coelicolor* cultivations

The contents of this chapter has been published in Petersen N, Ödman P, Cervera AE, Stocks S, Eliasson Lantz A, Gernaey KV. 2010. In situ NIR spectroscopy for analyte specific monitoring of glucose and ammonium in *Streptomyces coelicolor* fermentations. Biotechnology Progress 26: 263-271.

Abstract

There are many challenges associated with in situ collection of Near Infrared (NIR) spectra in a fermentation broth, particularly for highly aerated and agitated fermentations with filamentous organisms. In this study, antibiotic fermentation by the filamentous bacterium *Streptomyces coelicolor* was used as a model process. Partial Least Squares (PLS) regression models were calibrated for glucose and ammonium based on NIR spectra collected in situ. To ensure that the models were calibrated based on analyte specific information, semi-synthetic samples were used for model calibration in addition to data from standard batches. Thereby, part of the inherent correlation between the analytes could be eliminated. The set of semi-synthetic samples were generated from fermentation broth from five separate fermentations to which different amounts of glucose, ammonium, and biomass were added. This method has previously been used off-line but never before in situ. The use of semi-synthetic samples along with validation on an independent batch provided a critical and realistic evaluation of analyte specific models based on in situ NIR spectroscopy. The prediction of glucose was highly satisfactory resulting in a RMSEP of 1.1 g/L. The prediction of ammonium based on NIR spectra collected in situ was not satisfactory. A comparison with models calibrated based on NIR spectra collected off-line suggested that this is caused by signal attenuation in the optical fibers in the region above 2000 nm; a region which contains important absorption bands for ammonium. For improved predictions of ammonium in situ, it is suggested to focus efforts on enhancing the signal in that particular region.

8.1 Introduction

Fermentation processes are widely used in the production of food and food ingredients, pharmaceuticals, enzymes, and a number of bulk chemicals. Efficient

control of fermentation processes is often essential for obtaining a high yield, and relies on timely and accurate information concerning the physical, chemical, and biological conditions in the bioreactor. Today physical and chemical parameters such as for example the pH, temperature, dissolved oxygen concentration, and carbon dioxide evolution rate are measured reliably on-line whereas information concerning biologically important variables such as the concentration of different nutrients, metabolites, and the biomass is mainly available via laborious off-line analyses (Schügerl, 2001). Over the past decades there has been a great effort in developing methods for real-time monitoring of fermentation processes using various advanced sensors. The formulation of the Process Analytical Technology (PAT) guidance, launched by the U.S. Food and Drug Administration in 2004 (http://www.fda.gov/Cder/OPS/PAT.htm) has resulted in an increased activity in this research area.

Near Infrared (NIR) spectroscopy has a large potential for monitoring fermentation processes and several studies have been published on the subject (Arnold et al., 2000; Cavinato et al., 1990; Holm-Nielsen et al., 2008; Riley et al., 1998b; Rodrigues et al., 2008; Scarff et al., 2006; Vaidyanathan et al., 2001c). Most of the published studies discuss off-line and at-line NIR spectroscopy applications and can report low prediction errors for many nutrients such as glucose, glycerol, ammonium as well as biomass and various products (Crowley et al., 2005; Finn et al., 2006; Hall et al., 1996; Vaidyanathan et al., 2003). However, if the aim is automatic control, the measurements should be made on-line. On-line collection of NIR spectra poses a challenge as it does not allow for any sample pretreatment e.g. the separation of biomass and the broth. Different on-line setups for collection of NIR spectra in fermentation processes have been tested and reported in the literature such as a by-pass loop (Holm-Nielsen et al., 2008), a reflectance probe outside a glass window (Cavinato et al., 1990; Ge et al., 1994), and reflectance and transmission probes immersed in the fermentation broth (Arnold et al., 2003, 2002; Blanco et al., 2006; Jørgensen et al., 2004; Lewis et al., 2000; Navrátil et al., 2005). Probes placed inside the fermenter, here referred to as in situ setups, are subject to disturbances from agitation, air bubbles, solid particles (e.g. biomass or insoluble medium components), temperature, pH changes etc. Despite the large number of potential disturbances, successful studies have been reported for different cell cultures with probes placed inside the fermenter (Arnold et al., 2003; Lewis et al., 2000; Roychoudhury et al., 2007). In these papers, the processes under investigation are typically characterized by a low agitation rate and a low biomass concentration, which make these systems attractive for in situ NIR monitoring. In situ monitoring of various unicellular microbial systems such as yeast and *E. coli* run under more "challenging" conditions, i.e. high aeration and stirring rates, have also been reported in the literature (Arnold et al., 2002; Navrátil et al., 2005; Tamburini et al., 2003; Tosi et al., 2003). From a comparative study of in situ and at-line measurements of biomass in an *E. coli* cultivation, Arnold et al. (2002) have concluded on a number of challenges with in situ monitoring such as loss of wavelength regions (potentially above 2000 nm) due to adverse signal to noise ratio, loss of light intensity, gas phase effects, and mechanical vibration of equipment.

Filamentous organisms are widely used in the industry for the production of

various products such as enzymes, antibiotics, organic acids and flavors (Ratledge and Kristiansen, 2006). Studies concerning on-line monitoring applications on these systems are therefore of great interest. The filamentous organisms create a highly complex matrix for in situ spectroscopic measurements. Firstly, the filamentous growth results in a highly viscous fermentation broth exhibiting several non-Newtonian characteristics (Arnold et al., 2000). Secondly, the morphology often changes throughout the fermentation, which affects the scattering properties of the fermentation broth as well as the specific absorption of the biomass (Vaidyanathan et al., 2003). Finally, the filamentous organisms often have high demands for aeration, which further complicates in situ collection of NIR spectra. Only very little has been published so far about in situ monitoring of complex filamentous systems. Rodrigues et al. (2008) recently published a study concerning in situ monitoring of industrial scale *Streptomyces clavuligerus* fermentations. Good results were obtained for monitoring of the carbon source concentration, the viscosity, and product concentration (an active pharmaceutical ingredient). However, the problem of correlation between analytes was not considered in that study (see below).

A major challenge in the calibration of chemometric models for NIR monitoring of fermentations is the inherent correlation between the different analytes. A typical batch process will start with low biomass and product concentrations and high nutrient concentrations. As the fermentation progresses, the biomass and product concentrations will increase while the nutrients are taken up. This results in negative correlations between biomass and nutrients concentrations and positive correlations between the concentrations of different nutrients. If concentrations of nutrients, biomass and product are highly correlated it is possible to calibrate chemometric models for prediction of both nutrients and product based for example on the signal from biomass alone, so called correlated models. This may be acceptable if the yields are constant during the fermentation. However, in abnormal situations – or actually as soon as the yields change – the predictions will not be accurate. Different solutions have been proposed in the literature to overcome this problem, including: (a) the use of purely synthetic samples; (b) semi-synthetic samples (fermentation samples that have been altered by addition of different compounds of interest); (c) samples from multiple fermentations in which for example the media composition is changed to introduce variations between batches; and (d) combinations of the above (Blanco et al., 2004; Cervera et al., 2009; Finn et al., 2006; Lewis et al., 2000; McShane and Coté, 1998; Rhiel et al., 2002; Riley et al., 1998a,b).

There is a general agreement that calibration models based solely on synthetic samples perform poorly for the prediction of cultivation samples (Blanco et al., 2004; Lewis et al., 2000; McShane and Coté, 1998; Riley et al., 1998a). Riley et al. (1998b) have proposed an adaptive calibration scheme in which samples taken from the bioreactor at different time points were subsequently altered by adding different amounts of the analytes of interest, thereby breaking the correlations between them. This scheme has been used successfully in off-line studies by Rhiel et al. (2002) for the prediction of various analytes in cell culture medium and by Finn et al. (2006) for the prediction of glucose in yeast fermentations.

The purpose of the present study was to critically evaluate the ability of NIR

spectroscopy to measure important analytes in low concentrations in situ in a filamentous fermentation broth. Addressing the problem of correlations between analytes in the model building in situ was a central part of the evaluation since results cannot be transferred directly from the off-line setting. This study used a combination of standard batch samples and semi-synthetic samples to break the correlations between analytes and calibrate *analyte specific* models for glucose and ammonium. To highlight the specific challenges of collection of NIR spectra in situ, the samples were also measured using off-line NIR spectroscopy.

The results section begins with a prestudy performed to quantify different sources of variation *in situ* in an aerated and agitated fermenter. After this, the results of model calibration for prediction of glucose and ammonium are presented. In practice the collected measurements, both the NIR spectra as well as the off-line measurements are subjected to random noise. This noise will introduce uncertainty into the estimated models and the model performance. To quantify the uncertainty of the model performance it was chosen to repeat model calibration, optimization and validation for different permutations of the data. The results of this are presented at the end of the results section. Finally, the results are critically discussed

8.2 Materials and methods

8.2.1 Generation of semi-synthetic samples

The procedure for the generation of the semi-synthetic samples is illustrated in Figure 8.1. Five *Streptomyces coelicolor* batch fermentations were run in parallel. The five batches were harvested at different time points throughout the fermentation process, such that each of the different phases of the fermentation was represented by at least one batch. For each batch the broth was centrifuged after harvesting, which resulted in a concentrated biomass suspension and a supernatant. The major part of the supernatant was transferred to the fermenter with the in situ NIR probe. The remaining part of the supernatant was used to produce solutions with higher concentrations of either glucose or ammonium. The supernatant was diluted approximately 1:2 with distilled water prior to the addition of glucose/ammonium. Thus when adding the concentrated solutions of glucose/ammonium, the other analytes were slightly diluted thereby minimizing distortion towards higher concentrations in the final set of samples (Rhiel et al., 2002). The concentrated biomass suspension and the two solutions were added to the broth in the NIR-fermenter in a series of small steps - approximately 8 x 100 mL for each of the solutions. The biomass suspension was always added first followed by addition of the glucose solution and then the ammonium solution or vice versa. After each addition a set of NIR spectra were collected and a volume corresponding to the added volume was removed for off-line analysis of biomass, glucose, and ammonium concentrations.

8.2.2 Media

2YT preculture medium: 16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, pH 6.8. *Production medium*: 3 mM NaH_2PO_4, 100 mM NH_4Cl, 10 mM KCl, 2 mM

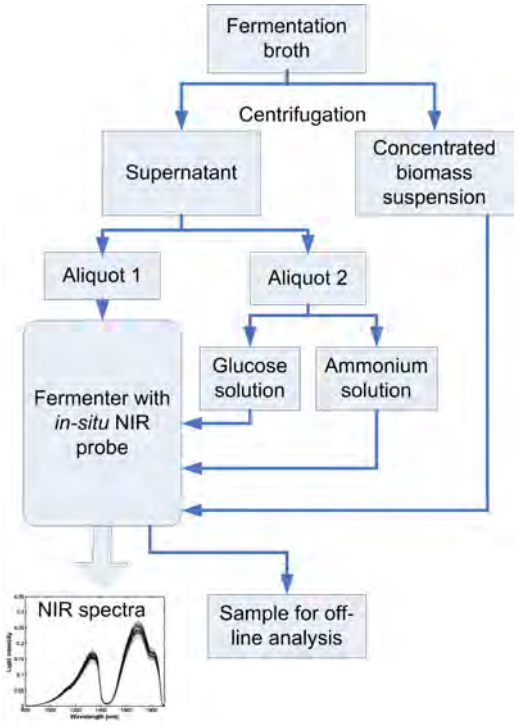


Figure 8.1: Experimental procedure for the generation of the semi-synthetic samples. See text for explanation.

Na_2SO_4, 2 mM citric acid, 1.25 mM $MgCl_2$, 0.25 mM $CaCl_2$, glucose 30 g/L, pH 6.9. The production medium was supplemented with 5 mL/L trace element solution (20 mM $FeCl_3$, 10 mM $CuCl_2$, 50 mM $ZnCl_2$, 10 mM $MnCl_2$, 0.02 mM Na_2MoO_4, 20 mM $CoCl_2$, 10 mM H_3BO_4) and 1 mL/L vitamin solution (Verduyn et al., 1992).

8.2.3 Strain and culture conditions

100 μL frozen mycelium of *Streptomyces coelicolor* M145 in 20% peptone (OD_{450} = 26) was inoculated into 50 mL 2YT preculture medium in a 500 mL baffled shake flask containing 30 glass beads (diameter = 3 mm). The shake flask was incubated for 24 h on a rotary shaker at 150 rpm and 30°C. The preculture was harvested when the pH had reached 7.7, corresponding to the late exponential growth phase. The culture, including glass beads, was transferred to a 50 mL plastic tube and centrifuged at 5000 rpm for 5 min. The medium was decanted, the cells were vortexed together with the glass beads for 10 seconds before re-suspension in production medium and concomitant inoculation of the bioreactor. Each bioreactor was inoculated with cells from two preculture flasks. All fermentations were carried out at 30°C, 600 rpm agitation and 1 vvm aeration. pH was kept constant at 6.9 by addition of 2 M NaOH. The parallel cultivations used for generation of the semi-synthetic samples were carried out in five in-house built 4 L glass fermenters. Two standard batches were carried out in a 3 L stainless steel bioreactor (MBR Bioreactor AG, Switzerland).

8.2.4 Analytical techniques

Glucose was determined on an HPLC system (Dionex Corporation, Sunnyvale, CA) equipped with an Aminex HPX-87H column (Bio-Rad Laboratories, Hercules, CA; operated at 60°C) and a refractometric detector (Shimadzu, Japan), using 5 mM H_2SO_4 as eluent with a flow rate of 0.6 mL/min.

Ammonium was determined using an ammonia-selective electrode (Metrohm, Switzerland).

Dry cell weight was determined by filtering a known amount of culture broth through a pre-weighted 0.45μm pore size filter (Sartorius, Germany). The filter was dried in 15 minutes at 150 W in a microwave oven before the weight gain was measured.

In situ NIR

in situ NIR spectra were collected with an ABB Bomem FTPA2000-260 spectrometer (ABB Bomem, Switzerland) equipped with a transflectance probe with a slit width of 0.5 mm corresponding to an optical path length of 1 mm submerged in the fermenter. The spectral data for each sample were collected as single scan spectra and averaged subsequently based on a minimum of 80 single scan spectra for each sample. The spectra were referenced against an air spectrum collected before the measurements of each batch. The broth was scanned over a range from 4000 to 15784 cm^{-1} (633 to 2500 nm) with a resolution of 32 cm^{-1}.

Off-line NIR

Off-line NIR spectra were collected with a FOSS NIRSystems spectrophotometer 6500 (FOSS NIRSystems Inc., MD, USA). The samples for off-line analysis were filtrated using a 0.45 μm filter and presented in a 1 mm cuvette in a transport module at room temperature. A ceramic reference spectrum was collected before each measurement. The resolution was 2 nm and the collected spectra were averages of 64 internal scans. The measurements were made in the range 400-2500 nm.

8.2.5 Model development and validation

Separate PLS models were calibrated for the prediction of glucose and ammonium based on the NIR spectra collected in situ and off-line. Data from seven different batches were available (five batches used to generate the semi-synthetic samples and two full length batches). A PLS model was calibrated based on the semi-synthetic samples (SS-1:5) and the standard batch 1 (ST-1). The wavelength interval, the number of latent variables (LV), and the pretreatment method were chosen in an automated calibration procedure by minimizing the Root Mean Square Error (RMSE) of leave-one-batch-out cross-validation. The calibrated model was subsequently validated on the standard batch 2 (ST-2).

The optimal wavelength interval was found among the wavelength intervals defined by all different combinations of the upper and lower bounds shown below.

- Lower bounds: 1100, 1200, 1300, 1350, 1400, 1500 nm.

- Upper bounds: 1700, 1800, 1850, 1900, 2000, 2100, 2200 nm.

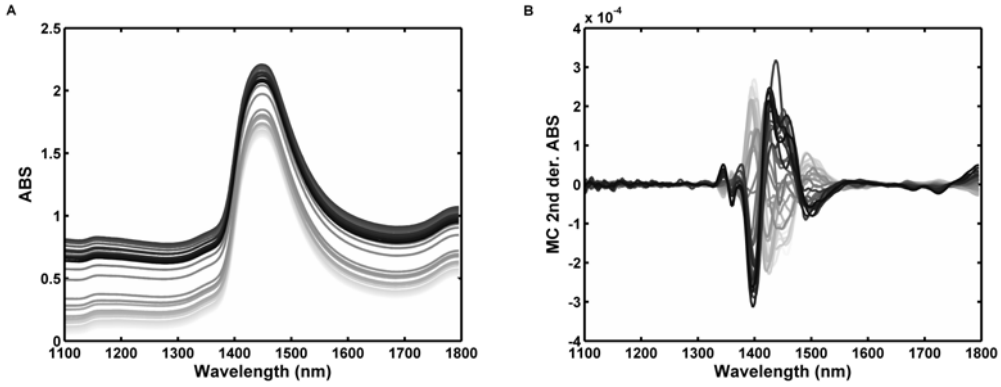
The wavelength regions below 1100 nm and above 2200 nm were excluded because a prestudy of the data had shown that these regions did not contribute positively to the quality of the models (results not included in this manuscript). In addition intervals in which the water peak between 1900 nm to 2000 nm was excluded were tested. Different pretreatment strategies were tested including 1^{st} derivative (Savitzky-Golay smoothing using 11, 13 and 15 points), 2^{nd} derivative (Savitzky-Golay smoothing using 11, 13 and 15 points), multiplicative scatter correction (MSC), standard normal variate (SNV), and a combination of SNV and the 1^{st} derivative methods. All model development was carried out using MATLAB 7 (Mathworks Inc, USA) and the PLS toolbox v. 4.2.1 from Eigenvector Inc. (USA).

8.3 Results

8.3.1 Prestudy

in situ NIR spectroscopy is subjected to additional sources of variation compared to the off-line setup. First of all the analyzed sample is heavily influenced by the environment in the fermenter and secondly the in situ setup is somewhat variable due to the connection via fiber optic cables. A prestudy was performed to evaluate the effect on the spectra of these sources of variation.

In the *Streptomyces coelicolor* fermentations the morphology changed from large pellets to small loose clumps. Differences in morphology were also observed between each batch. These differences caused variations in the NIR spectra. The biomass particles scattered the light and caused a baseline shift in the absorbance spectra (Figure 8.2a). The scattering properties depend on the amount and size of the particles. As the main interest in this study was the concentration of glucose and ammonium, the scattering should preferably be removed. Since removing biomass was not an option in situ, the possibility to remove the scatter effect by pre-treating the spectra was evaluated. The second derivative pretreatment of the spectra largely removed the scattering. However, some time dependent differences were still present between the spectra. The differences were largest around the water peak but were found throughout the NIR spectra (Figure 8.2b). These differences were partly attributed to the changing biomass concentration since the effect of biomass is not solely a baseline shift caused by scattering. Firstly, the presence of dry matter displaces the water thus affecting the water peak. Secondly, biomass itself is absorbing in the NIR region (Vaidyanathan et al., 2001b, 1999b). The many time dependent changes in the NIR spectra made it very difficult to distinguish the small absorption peaks of analytes present in low concentrations. Multivariable modeling tools, such as PLS regression, are capable of identifying and using small differences in the spectra not immediately visible to the eye. However, in order to guide the model to these unique absorption peaks, the data set should include samples in which the analytes of interest are uncorrelated to other components affecting the NIR spectra. This was achieved by designing an experiment in which semi-synthetic samples were measured in situ.



(a) Raw absorbance spectra **(b)** 2^{nd} derivative absorbance spectra

Figure 8.2: NIR spectra collected in situ in standard batch 1. (a) Raw absorbance spectra; (a) Mean centered 2^{nd} derivative absorbance spectra. The spectra are colored according to the sample number ranging from gray (start of the batch) to black (end of the batch).

The common practice for collection of NIR spectra is to collect an average spectrum calculated based on a number of single scans normally ranging between 32 to 128 (Arnold et al., 2002; Rodrigues et al., 2008). Nonetheless, it is interesting for the in situ NIR setup to study how much the single scan spectra vary due to the rapidly changing physicochemical conditions around the probe. In an agitated fermentation broth with both air bubbles and differently sized mycelium particles, the single scans will inevitably differ depending on the nature of the sample passing by the measurement window at the time of the measurement. A large variation was observed between single scans in the fermentation broth at normal operating conditions (1 vvm aeration and 600 rpm stirring). Tests of the effect of stirring and aeration showed that a decrease in the stirring rate (from 600 to 300 rpm) led to a dramatic reduction in the variation between the single spectra as well as a small decrease in the average absorbance. The effect of aeration was found to be less pronounced compared to the effect of the stirring rate, which is consistent with the results obtained in previous studies (Navrátil et al., 2005; Tamburini et al., 2003). After pretreatment of the spectra (2^{nd} derivative) the effect of both stirring and aeration was small (about one fifth in magnitude) compared to the within batch variation and hence these effects were not considered further.

In addition to the effects of the turbulent environment inside the fermenter, it was found that small changes in the in situ NIR instrument setup, such as differences in the bending of the optic cables and connection of the optical cables to the probe, resulted in significant changes between the collected spectra. A test was performed by collecting pure water spectra interrupted by a detachment and a reattachment of the fiber optic cables to the in situ probe. The results showed that the variation between the pretreated spectra caused by movement and reattachment of the cables was in the same order of magnitude as within batch variation. This posed a challenge for the model development since the model

should be able to distinguish between spectral changes caused by the experimental setup and changes caused by differences in the analyte concentrations. Due to the fragile nature of the fiber optic cables, it was necessary to disconnect the probes during autoclavation of the fermenters. Consequently, each batch represented a different setup. In order to calibrate models that were robust against variations in the setup it was therefore chosen to use leave-one-batch-out cross-validation in the development of prediction models. Furthermore, in order to provide a realistic estimate of the predictive ability of the resulting model, it was chosen to validate the model on a new and independent batch.

8.3.2 Data set for model development

A total of seven batch fermentations were run; five batches (SS-1:5) for the generation of the set of semi-synthetic samples as well as two standard batches (ST-1:2). This resulted in a data set consisting of 108 semi-synthetic samples including broth from five batches (SS-1:5) as well as an additional 34 and 30 samples collected during two standard batches, ST-1 and ST-2, respectively. In situ as well as off-line NIR spectroscopic measurements were collected for all samples. The correlation coefficient between glucose and ammonium in the samples from the two standard batches was found to be 0.93, which signified a high degree of correlation (1 is perfect positive correlation). In the combined set of samples, the correlation coefficient between glucose and ammonium was reduced to 0.76. Similarly, the correlation between biomass and the two nutrients, glucose and ammonium, was reduced from -0.92 and -0.97 to -0.76 and -0.79 respectively in the combined set of samples. Such a reduction of the correlation coefficients between analytes was precisely the goal of the inclusion of the semi-synthetic samples.

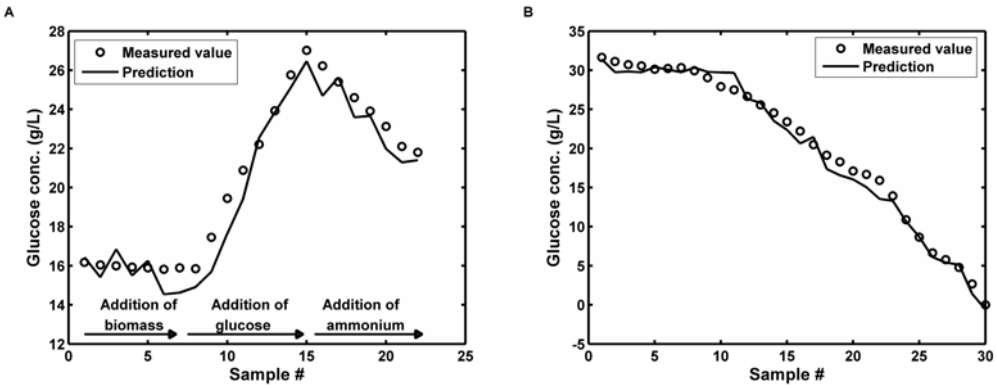
8.3.3 Prediction of glucose

PLS models were calibrated for the prediction of glucose based on the NIR spectra collected in situ and off-line. The five sets of semi-synthetic samples (SS-1:5) and the standard batch 1 (ST-1) were used for model calibration. The pre-treatment, wavelength region, and the number of Latent Variables (LV) were determined by minimizing the Root Mean Square Error of leave-one-batch-out cross-validation (RMSECV). Following this, an independent data set (standard batch 2, ST-2) was used to validate the model thus providing the Root Mean Square Error of Prediction (RMSEP).

The RMSECV and the RMSEP for glucose were 1.9 g/L and 1.1 g/L respectively for the in situ model compared to 1.4 and 1.4 respectively for the off-line model (Table 8.1). Hence, the cross-validation and the validation errors were low and in the same range for both the off-line and the in situ models. The prediction error of 1.1 g/L achieved by the in situ model corresponds to 5.4% of the average glucose concentration in the validation set, which is in the same range as the reference method. Figure 8.3 shows the cross-validation predictions for SS-3 (a) and validation predictions for ST-2 (b) for the in situ model. For the semi-synthetic samples (Figure 8.3a) the changes in glucose concentration were very well predicted, both during addition of biomass, ammonium, and glucose, indicating that the models were indeed based on a signal from the glucose. Likewise, the plot

Table 8.1: Results of PLS model calibration and validation for the prediction of glucose based on the NIR spectra collected off-line and in situ.

Spectra	Wave-lengths	Pre-treatment	LVs	RMSECV (0-43 g/L)	RMSEP (0-32g/L)	RMSEP (% of mean value)
in situ	1100-1900	2^{nd} der.	16	1.9	1.1	5.4
Off-line	1400-1800	SNV + 1^{st} der.	10	1.4	1.4	6.8



(a) Cross-validation **(b)** Validation

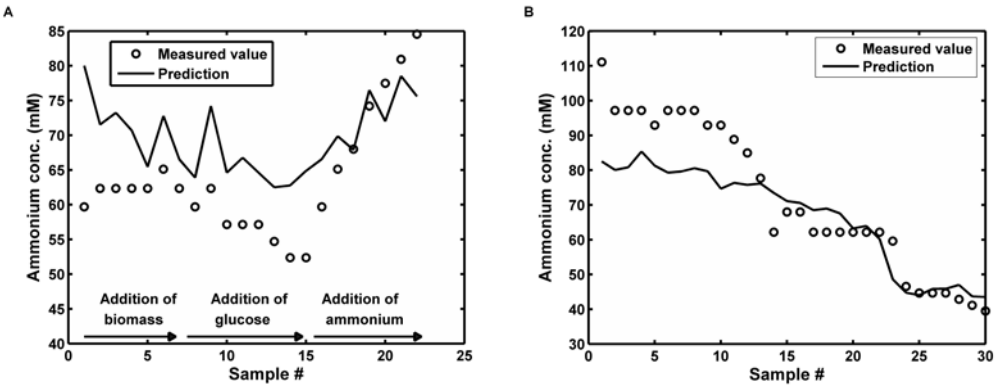
Figure 8.3: Prediction of the glucose concentration using the model calibrated based on the in situ spectra. (a) cross-validation results for semi-synthetic sample set 3 (SS-3); (b) validation results for the independent standard batch 2 (ST-2).

of the predictions of ST-2 (Figure 8.3b) shows that the model was also able to capture the variation in the glucose concentration throughout a standard batch. Overall the model performed highly satisfactory.

Using the automated calibration procedure (see section 8.2), a relatively high number of LVs (Blanco et al., 2006) was found to be optimal for prediction of glucose in situ. A closer look at the RMSECV curve showed that the number of LVs could be reduced to 14 without a significant increase in RMSECV (RMSECV = 1.96 g/L and RMSEP = 1.42 g/L). 14 is still a high number of LVs which entails a danger of overfitting the data. Nonetheless, it was chosen to accept this high number of LVs as the leave-one-batch-out cross-validation provides a very tough test of the model that should ensure robustness across the batches. Moreover, the external validation using the last standard batch, which was independent of all model optimization, provided the final test that the data had not been overfitted. For the off-line model the automated calibration procedure selected 10 LVs (Table 8.1). Using manual selection of the number of LVs, it was possible to reduce the number of LVs to six without a significant increase in the RMSECV (RMSECV = 1.63 g/L and RMSEP = 1.57 g/L). The off-line data thus could be modeled with a lower number of components compared to the in situ data.

Table 8.2: Results of PLS model calibration and validation for the prediction of ammonium based on the NIR spectra collected off-line and in situ.

Spectra	Wave-lengths	Pre-treatment	LVs	RMSECV (52-109mM)	RMSEP (39-111mM)	RMSEP (% of mean value)
in situ	1500-1850	SNV + 1st der.	6	7.1	11	16
Off-line	1200-1850, 2000-2200	SNV + 1st der.	7	3.6	5.1	7.2



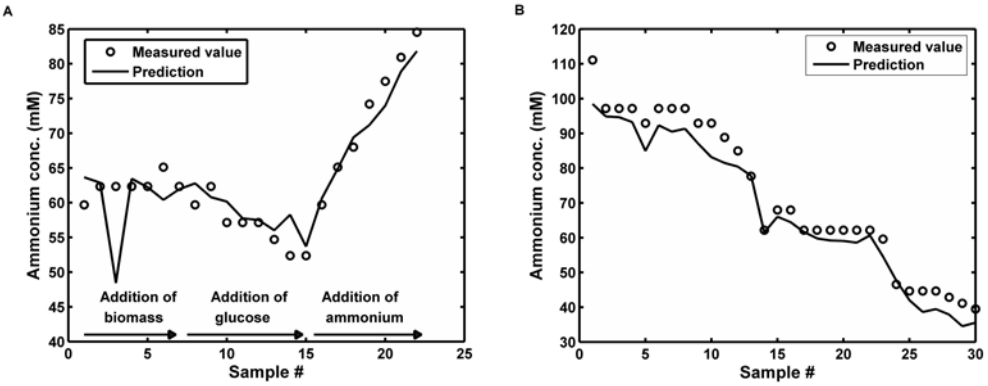
(a) Cross-validation **(b)** Validation

Figure 8.4: Prediction of the ammonium concentration using the model calibrated based on the in situ spectra.(a) cross-validation results for semi-synthetic sample set 3 (SS-3); (a) validation results for the independent standard batch 2 (ST-2).

8.3.4 Prediction of ammonium

Models for prediction of ammonium were calibrated and validated following the same procedure as for glucose. The resulting RMSECV and the RMSEP for the in situ model were 7.1 mM and 11 mM respectively (Table 8.2). This corresponds to a prediction error of 16 % of the mean value in the validation set, which is relatively high. Figure 8.4 shows the cross-validation predictions for SS-3 (a) and validation predictions for ST-2 (b) of ammonium using the in situ model. The changes in ammonium concentration in the set of semi-synthetic samples and the standard batch were only partly captured by the model. Furthermore, the predictions of the standard batch were highly biased for the first 12 samples. Hence, the performance of the in situ model for prediction of ammonium was not satisfactory.

For comparison, the ammonium prediction errors for the model based on spectra collected off-line were approximately half of the errors for the in situ model (Table 8.2). The off-line model was able to capture the variation of the ammonium concentration in both the semi-synthetic as well as the standard samples



(a) Cross-validation **(b)** Validation

Figure 8.5: Prediction of the ammonium concentration using the model calibrated based on the off-line spectra. (a) cross-validation results for semi-synthetic sample set 3 (SS-3); (b) Validation results for the independent standard batch 2 (ST-2).

with the exception of one slightly outlying sample (Figure 8.5).

8.3.5 Repetition of model calibration and validation for quality control

The model building procedure was repeated six more times; each time using one of the previously used calibration batches (SS-1:5 and ST-1) for external validation and the remaining six (including ST-2) for model calibration and optimization. This was done to verify the robustness of the data and the model building procedure. The different partitions of the data and subsequent model calibration and optimization resulted in six new models in addition to the above presented model. The mean value and the standard deviation of the RMSECV and the RMSEP for the in situ and off-line models are listed in Table 8.3. For the in situ glucose model both the RMSECV and RMSEP were in the same range for all seven partitions of the data. This resulted in average values of 1.9 ± 0.28 and 2.0 ± 1.1 g/L for the RMSECV and RMSEP respectively. Small differences in the prediction errors were mainly caused by differences in the absolute values of the different validation sets and therefore the average value and the standard deviation of the relative RMSEP were low ($8.7 \pm 2.6\%$). This indicated that the models were robust. For glucose the prediction errors for the in situ and the off-line models were in the same range. For ammonium, however, the relative RMSEP was $15 \pm 9.4\%$ for the in situ model compared to $5.4 \pm 1.5\%$ for the off-line model. This is a substantial difference and thus verifies the observations made for the first prediction models. The high standard deviation of the prediction errors for the in situ ammonium models indicated some potential problems of the validity of the calibrated model for the prediction of new and independent batches.

In this study all of the off-line models for prediction of ammonium were utilizing the spectrum up to 2200 nm, whereas only one of the in situ models included the

Table 8.3: Results of repeated model calibration, optimization and validation with the seven different partitions of the data set. The RMSEP is shown in original units (g/L and mM for glucose and ammonium respectively) and as percentage of the average value.

	Glucose mean ± std			Ammonium mean ± std		
Spectra	RM-SECV (g/L)	RMSEP (g/L)	RMSEP (%)	RM-SECV (mM)	RMSEP (mM)	RMSEP (%)
in situ	1.9±0.28	2.0±1.1	8.7±2.6	7.1±0.96	11±7.6	15±9.4
Off-line	1.2±0.16	1.5±0.5	7.1±1.3	3.5±0.27	4.0±1.2	5.4±1.5
Off-line (1100-1900 nm)	-	-	-	4.3±0.59	6.5±5.3	8.8±7.4

region above 1900 nm (data not shown). For comparison, models were calibrated for the prediction of ammonium based on the off-line spectra including only the region below 1900 nm. The results are summarized in Table 8.3. There was an average increase in the prediction error of 2.5 mM corresponding to an increase of 63 %, when excluding the region above 1900 nm.

8.4 Discussion

To ensure analyte specificity of the PLS models, it was crucial that the inherent correlation between biomass, glucose, and ammonium was reduced. This was achieved by measuring a large number of semi-synthetic samples in the fermenter using in situ NIR spectroscopy. The inclusion of semi-synthetic samples successfully reduced the correlations between ammonium, glucose, and biomass from ±0.92-0.97 to ±0.77-0.79. This study thus presents an example of a strategy for implementation of NIR monitoring on a fermentation process by the use of semi-synthetic samples in situ along with a thorough calibration and validation procedure, resulting in a critical and realistic estimate of the performance of NIR spectroscopy in situ. We believe that such critical evaluations are necessary for the further development of NIR spectroscopy as part of PAT applications in the biotechnological industry.

The results showed that the glucose concentration could be predicted satisfactorily based on NIR spectra collected in situ. The prediction error for the in situ model was 1.1 g/L corresponding to 5.4% of the average value in the validation set. In comparison Rodrigues et al. (2008) achieved a relative error of 13% for the carbon source in the only other published study concerning in situ NIR measurements in filamentous fermentations. The plots of the predicted glucose concentrations in both the semi-synthetic and the standard samples showed good predictions regardless of the concurrent increase or decrease in concentration of the other analytes in solution (Figure 8.3). These plots thus verified model specificity and constituted a visual check of the quality of the calibrated models.

Furthermore, the repeated model calibration and validation resulted in a relatively low standard deviation of the prediction error for glucose. This increased certainty that the obtained results are generalizable. Real-time predictions of glucose will provide valuable insight into many fermentation processes and can for example be used for automatic control of glucose addition in fed-batch processes. Both process knowledge and implementation of automatic control to keep a process within its design space are important aspects of PAT.

For ammonium the validation of the in situ model resulted in a relative RMSEP of 16 % corresponding to a correlation coefficient of 0.73, which is very similar to the correlation coefficient of 0.716 obtained by Rodrigues et al. (Rodrigues et al., 2008). However, this is not considered satisfactory in this study – particularly since the correlation coefficient between biomass and ammonium was found to be 0.79 in the present data set, indicating that it was not possible to construct an analyte specific in situ model for ammonium. The repeated model calibration (Table 8.3) showed that both the average value as well as the standard deviation of the prediction errors was higher for the in situ models compared to the off-line models. Comparisons to off-line models calibrated using either the full spectrum or the spectrum up to 1900 nm indicated that part of the explanation for the poor quality of the in situ models may be signal attenuation in the optical fibers above 2000 nm. This complies well with previous observations (Arnold et al., 2002). Ammonium has previously been predicted successfully in a number of studies using NIR spectroscopy. A literature survey revealed that the majority of the identified absorption peaks for ammonium are found above 2100 nm (Arnold et al., 2000, 2003; Chung et al., 1996; Rhiel et al., 2002; Vaidyanathan et al., 1999a). Few studies report absorption peaks around 1680-1690 nm (Vaidyanathan et al., 2001a,b) and at 900 nm (Vaidyanathan et al., 2001a). However, these peaks are placed within the overtone regions (first and third respectively) of the C-H stretch and cannot immediately be assigned to any of the theoretical absorption peaks of ammonium (Osborne et al., 1993). This obviously presents some challenges for in situ monitoring of ammonium, and highlights the importance of breaking inherent correlations when calibrating analyte specific prediction models. If monitoring of ammonium is required, it should be considered to optimize the signal in the region above 2100 nm. Using fluoride optical fibers instead of the standard glass fibers would be one way to improve the signal to noise ratio in that important wavelength region.

The calibration and validation of prediction models can be a time consuming and laborious task including data collection and the model building itself. Fermentation experiments require many resources and therefore the amount of experimental data available for building chemometric models will often be limited. The present study used data from a relatively low number of seven different batches. However, by implementing a procedure of repeating model calibration and validation it was possible to get an estimate of the uncertainty of the prediction error and thus increase confidence in the results. The chemometric models can be calibrated using different combinations of pretreatments, modeling algorithms, wavelength selections etc. The models in this study have been calibrated in a highly automated fashion using simple pretreatments, a rather crude wavelength selection, and standard PLS modeling. Such an automatic procedure will

be useful in an industrial setting where time for extensive model optimization will be limited. After a good model has been identified the model can be further optimized e.g. by manual selection of the number of components and fine tuning of the wavelength regions.

In spite of a number of studies on the subject, NIR spectroscopy is not a standard monitoring tool in fermentation processes. For NIR spectroscopy to be accepted in the biotechnological industry it is important to have guidelines for easy implementation like the automated procedure discussed above. Moreover, we believe it is essential that the analyte specificity is ensured and documented in on-line applications. To the best of our knowledge this study is the first to use semi-synthetic samples in situ. Models calibrated and validated based solely on standard batch samples may be highly useful for monitoring of a process in normal situations. However, if the process for some reason deviates from the normal trend, a model based on correlations is likely to have a poor predictive performance since the relations between components will be changed. For successful use outside normal operation, e.g. process monitoring or control purposes, it is important that the model is analyte specific. By including synthetic or semi-synthetic samples the variation in the data set is increased. This may increase the prediction errors but will improve model specificity by breaking the correlations, and will thus broaden the range of validity of a PLS model.

8.5 Conclusion

The main objective of the present study was to provide a critical and realistic evaluation of in situ NIR spectroscopy for measurement of glucose and ammonium in filamentous fermentations. This was achieved by using semi-synthetic samples in situ to break correlations and ensure analyte specificity and by validating the resulting model on an independent batch. The results showed that glucose could be predicted satisfactorily based on in situ NIR spectroscopy. On-line prediction of glucose will be of high value in many fermentation processes for example to control the glucose addition in fed-batch processes. The models calibrated for ammonium based on in situ NIR were not satisfactory. This may be partly explained by signal attenuation in optical fibers above 2000 nm. Thus, if monitoring of ammonium is important, it is suggested to improve the signal in this region.

Acknowledgments

Acknowledgement is made of Lisbeth Tina Hansen and Frans van den Berg at the Faculty of Life Sciences of the University of Copenhagen for the use of the off-line NIR spectrophotometer. The Ph.D. projects of Nanna Petersen and Peter Ödman are supported by a grant from the Innovative Bioprocess Technology Research Consortium financed by the Danish Research Council for Technology and Production Sciences, Chr. Hansen A/S, Danisco A/S and Novozymes A/S.

Bibliography

Arnold SA, Crowley J, Vaidyanathan S, Matheson L, Mohan P, Hall J, Harvey LM, McNeil B. 2000. At-line monitoring of a submerged filamentous bacterial cultivation using near-infrared spectroscopy. Enzyme and Microbial Technology 27:691–697.

Arnold SA, Crowley J, Woods N, Harvey LM, McNeil B. 2003. In-situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation. Biotechnology and Bioengineering 84:13.

Arnold SA, Gaensakoo R, Harvey LM, McNeil B. 2002. Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch *Escherichia coli* process. Biotechnology and Bioengineering 80:405–413.

Blanco M, Peinado AC, Mas J. 2004. Analytical monitoring of alcoholic fermentation using NIR spectroscopy. Biotechnology and Bioengineering 88:536.

Blanco M, Peinado AC, Mas J. 2006. Monitoring alcoholic fermentation by joint use of soft and hard modelling methods. Analytica Chimica Acta 556:364–373.

Cavinato AG, Mayes DM, Ge Z, Callis JB. 1990. Noninvasive method for monitoring ethanol in fermentation processes using fiber-optic near-infrared spectroscopy. Analytical Chemistry 62:1977–1982.

Cervera AE, Petersen N, Gernaey KV, Lantz AE, Larsen A. 2009. Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. Biotechnology Progress 25:1561–1581.

Chung H, Arnold MA, Rhiel M, Murhammer DW. 1996. Simultaneous measurements of glucose, glutamine, ammonia, lactate, and glutamate in aqueous solutions by near-infrared spectroscopy. Applied Spectroscopy 50:270.

Crowley J, Arnold S, Wood N, Harvey L, McNeil B. 2005. Monitoring a high cell density recombinant *Pichia pastoris* fed-batch bioprocess using transmission and reflectance near infrared spectroscopy. Enzyme and Microbial Technology 36:621–628.

Finn B, Harvey LM, McNeil B. 2006. Near-infrared spectroscopic monitoring of biomass, glucose, ethanol and protein content in a high cell density baker's yeast fed-batch bioprocess. Yeast 23:507–517.

Ge Z, Cavinato A, Callis J. 1994. Noninvasive spectroscopy for monitoring cell density in a fermentation process. Analytical Chemistry 66:1354–1362.

Hall JW, McNeil B, Rollins MJ, Draper I, Thompson BG, Macaloney G. 1996. Near-infrared spectroscopic determination of acetate, ammonium, biomass, and glycerol in an industrial *Escherichia coli* fermentation. Applied Spectroscopy 50:102.

Holm-Nielsen JB, Lomborg CJ, Oleskowicz-Popiel P, Esbensen KH. 2008. Online near infrared monitoring of glycerol-boosted anaerobic digestion processes: Evaluation of process analytical technologies. Biotechnology and Bioengineering 99:302.

Jørgensen P, Pedersen JG, Jensen EP, Esbensen KH. 2004. On-line batch fermentation process monitoring (NIR)-introducing 'biological process time'. Journal of Chemometrics 18:81.

Lewis CB, McNichols RJ, Gowda A, Coté GL. 2000. Investigation of near-infrared spectroscopy for periodic determination of glucose in cell culture media in situ. Applied Spectroscopy 54:1453.

McShane MJ, Coté GL. 1998. Near-infrared spectroscopy for determination of glucose, lactate, and ammonia in cell culture media. Applied Spectroscopy 52:1073.

Navrátil M, Norberg A, Lembrén L, Mandenius C. 2005. On-line multi-analyzer monitoring of biomass, glucose and acetate for growth rate control of a *Vibrio cholerae* fed-batch cultivation. Journal of Biotechnology 115:67–79.

Osborne BG, Fearn T, Hindle P. 1993. Practical NIR spectroscopy with applications in Food and Beverage Analysis. Harlow, UK: Longman Scientific and Technical.

Ratledge C, Kristiansen B. 2006. Basic Biotechnology. Cambridge, UK: Cambridge University Press.

Rhiel M, Cohen MB, Murhammer DW, Arnold MA. 2002. Nondestructive near-infrared spectroscopic measurement of multiple analytes in undiluted samples of serum-based cell culture media. Biotechnology and Bioengineering 77:73–82.

Riley MR, Amold MA, Murhammer DW. 1998a. Matrix-enhanced calibration procedure for multivariate calibration models with near-infrared spectra. Applied Spectroscopy 52:1339–1347.

Riley MR, Arnold MA, Murhammer DW, Walls EL, DelaCruz N. 1998b. Adaptive calibration scheme for quantification of nutrients and byproducts in insect cell bioreactors by near-infrared spectroscopy. Biotechnology Progress 14:527–533.

Rodrigues LO, Vieira L, Cardoso JP, Menezes JC. 2008. The use of NIR as a multi-parametric in situ monitoring technique in filamentous fermentation systems. Talanta 75:1356–1361.

Roychoudhury P, Kennedy R, McNeil B, Harvey LM. 2007. Multiplexing fibre optic near infrared (NIR) spectroscopy as an emerging technology to monitor industrial bioprocesses. Analytica Chimica Acta 590:110–117.

Scarff M, Arnold S, Harvey L, McNeil B. 2006. Near infrared spectroscopy for bioprocess monitoring and control: Current status and future trends. Critical Reviews in Biotechnology 26:17–39.

Schügerl K. 2001. Progress in monitoring, modeling and control of bioprocesses during the last 20 years. Journal of biotechnology 85:149–173.

Tamburini E, Vaccari G, Tosi S, Trilli A. 2003. Near-infrared spectroscopy: a tool for monitoring submerged fermentation processes using an immersion optical-fiber probe. Applied Spectroscopy 57:132–138.

Tosi S, Rossi M, Tamburini E, Vaccari G, Amaretti A, Matteuzzi D. 2003. Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes. Biotechnology Progress 19:1816–1821.

Vaidyanathan S, Arnold SA, Matheson L, Mohan P, McNeil B, Harvey LM. 2001a. Assessment of near-infrared spectral information for rapid monitoring of bioprocess quality. Biotechnology and Bioengineering 74:376–388.

Vaidyanathan S, Harvey L, McNeil B. 2001b. Deconvolution of near-infrared spectral information for monitoring mycelial biomass and other key analytes in a submerged fungal bioprocess. Analytica Chimica Acta 428:41–59.

Vaidyanathan S, Macaloney G, Harvey L, McNeil B. 2001c. Assessment of the structure and predictive ability of models developed for monitoring key analytes in a submerged fungal bioprocess using near-infrared spectroscopy. Applied Spectroscopy 55:444–453.

Vaidyanathan S, Macaloney G, Vaughan J, McNeil B, Harvey LM. 1999a. Monitoring of submerged bioprocesses. Critical Reviews in Biotechnology 19:277–316.

Vaidyanathan S, McNeil B, Macaloney G. 1999b. Fundamental investigations on the near-infrared spectra of microbial biomass as applicable to bioprocess monitoring. The Analyst 124:157–162.

Vaidyanathan S, White S, Harvey LM, McNeil B. 2003. Influence of morphology on the near-infrared spectra of mycelial biomass and its implications in bioprocess monitoring. Biotechnology and Bioengineering 82:715–724.

Verduyn C, Postma E, Scheffers WA, van Dijken JP. 1992. Effect of benzoic acid on metabolic fluxes in yeast: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 8:501–517.

Chapter 9

Evaluation of on-line sensors for biomass measurement in filamentous cultivation process development

The contents of this chapter is in preparation for submission to Journal of Biotechnology in Rønnest NP, Stocks S, Eliasson Lantz A, and Gernaey KV. Evaluation of on-line sensors for biomass measurement in filamentous cultivation process development.

Abstract

The purpose of this study was firstly to compare the performance of various on-line biomass sensors, and secondly to demonstrate their use in early development of a filamentous cultivation process. Eight *Streptomyces coelicolor* fed-batch cultivations were run as part of a process development in which the pH, the feed mode, and the medium composition were varied. All the cultivations were monitored in situ using multi-wavelength fluorescence (MWF) spectroscopy, scanning dielectric spectroscopy (DE), and turbidity. In addition, we logged all of the classical cultivation data, such as the carbon dioxide evolution rate (CER) and the concentration of dissolved oxygen. Prediction models for the biomass concentrations were estimated based on the individual sensors and on combinations of the sensors. The results showed that the more advanced sensors based on MWF and scanning DE did not offer any advantages over the simpler sensors based on dual frequency DE, turbidity and CER measurements for prediction of biomass concentration. By combining CER, DE, and turbidity measurements, the prediction error was reduced to 1.5 g/L, corresponding to 6 % of the covered biomass range. Moreover, by using multiple sensors it was possible to check the quality of the individual predictions and switch between the sensors in real-time.

9.1 Introduction

Real-time measurement of biomass concentration could be very useful for many cultivation processes. However, despite the development of a number of promising technologies, accurate measurement of biomass concentration in real-time remains a significant challenge – particularly for highly aerated and agitated cultivations of organisms with complex morphology. The challenges of biomass measurement have been excellently reviewed by Olsson and Nielsen (1997), Kiviharju et al. (2008) and Madrid and Felice (2005). The on-line sensors typically ap-

plied for biomass concentration measurement include sensors based on dielectric spectroscopy (DE), software sensors based on classical process monitoring data, and optical sensors of several kinds: on-line turbidity probes, near-infrared (NIR) reflectance and transmission probes, and multi-wavelength fluorescence (MWF). This study evaluates the use of scanning and dual frequency DE spectroscopy, simple software sensors, turbidity, and MWF spectroscopy.

On-line turbidity probes are commercially available at a relatively low price. Turbidity probes measure the absorbance of light in a limited range, which may be in the visible or in the NIR region. Measurements in the NIR region offer some advantages since colored compounds do not interfere in this range. In a recent comparison, Kiviharju et al. (2007) concluded that a commercial turbidity probe measuring in the NIR range was superior to DE for the measurement of biomass in a range of bacterial and yeast cultivations. The experiments also included cultivations with the filamentous organism *Streptomyces peucetius*, but the results for the turbidity probe were not reported. The advantages of the turbidity probe include robustness, and the fact that optical density is an established off-line measure of biomass concentration. The main disadvantages are its limited range of linearity, its sensitivity to interference from air bubbles and particulate matter, and its dependence on cell morphology, which is particularly problematic in processes with filamentous organisms.

Dielectric spectroscopy is based on the ability of viable cells to store electrical charge. On-line, DE has mainly been used for the monitoring of biomass concentrations in cultivations with unicellular organisms (Arnoux et al., 2005; Dabros et al., 2009a; Markx et al., 1991a; November and Van Impe, 2000; Xiong et al., 2008). In filamentous systems, good correlations between biomass concentration and capacitance have been obtained on-line in the exponential, transition, and stationary growth phases (Neves et al., 2000; Sarra et al., 1996). However, different correlations between biomass and capacitance were observed in the biomass decline phase (Neves et al., 2000; Sarra et al., 1996). Recently, scanning DE has made it possible to obtain capacitance measurements over a range of frequencies, which may provide additional information about the size and distribution of cells. To our knowledge, this has not previously been applied to filamentous cultivations. The advantages of DE include a broad range of linearity, the selectivity to live biomass, and its insensitivity to other particulate matter (Harris et al., 1987). The main disadvantages are the effects of changes in the conductance of the medium and problems with polarization of the probe, causing noise in the measurements.

In multi-wavelength fluorescence (MWF) spectroscopy, a 2-dimensional landscape covering several emission and excitation wavelengths is measured, which makes it possible to quantify a number of biological fluorophores, such as proteins and cofactors in the cultivation broth. MWF spectroscopy has been used for monitoring of biomass in a number of studies including both unicellular and filamentous organisms (Clementschitsch et al., 2005; Haack et al., 2007; Ödman et al., 2010). The main advantages of MWF are its sensitivity and selectivity towards the fluorescent compounds. Furthermore, under the right conditions, MWF can provide additional information about the metabolic state of the organisms (Hantelmann et al., 2006). The use of MWF for the estimation of biomass

concentration depends on the correlation between the biological fluorophores and the biomass concentration, which may only be applicable under certain process conditions (Lantz et al., 2006; Li et al., 1991; Ödman et al., 2009).

Software sensors based on standard monitoring data, such as the carbon dioxide evolution rate (CER), the oxygen uptake rate (OUR), and the base consumption, have been developed for the estimation of biomass concentrations in a number of cultivation systems (Jenzsch et al., 2007; Ödman et al., 2009; Sundström and Enfors, 2008). Carbon dioxide evolution and base consumption are closely related to the growth of biomass, but the validity of the predictions is based on the assumption that the yields are constant or at least predictable. The advantages of the software sensors include their low price and the significant amount of experience and knowledge accumulated on how to analyze CER and OUR due to their established use for monitoring in the biotechnological industry.

The development of suitable on-line biomass sensors has been going on for decades, but as far as we know, most of the above-mentioned sensors have only found limited application in the bioprocess industry so far. This illustrates the difficulty of measuring the biomass concentration and the high requirements of the industry with respect to accuracy, robustness, and the range of applicability of the sensors. Different stages of industrial production, such as process development or large scale production, present different requirements for the sensors. The process development, with its high level of process variation, requires robust sensors that are quickly calibrated. So far, the majority of studies concerning on-line biomass sensors have focused on measurements in similar processes (Dabros et al., 2009a; Kiviharju et al., 2007; Markx et al., 1991b; Neves et al., 2000; November and Van Impe, 2000; Xiong et al., 2008). However, a number of previous studies have included some process variation, most commonly different carbon sources and different concentration levels (Arnoux et al., 2005; Haack et al., 2007; Jenzsch et al., 2006; Ödman et al., 2009; Sundström and Enfors, 2008).

The aim of this study was to test and compare some of the most prominent on-line biomass sensors in the development of a filamentous cultivation process. Firstly, the sensors were calibrated both individually and in combination in order to compare their performance. The results are presented based on all of the cultivations from the process development (eight batches). We then tested whether a selected biomass sensor could be calibrated and used for biomass estimation *during* the short process development. Furthermore, the biomass sensor was combined with a simple supervision strategy to explore whether this could improve the usability of the sensor.

9.2 Materials and methods

9.2.1 Overview of the cultivations

In total eight cultivations were run as part of a process development study in which the medium, pH and feed mode were varied (Table 9.1). In batches A1-4 the batch medium and the feed were solutions of glucose, casamino acids, salts and vitamins. In the feed, the nutrient concentrations were multiplied by a factor of 3 compared to the batch medium except for the trace metals and vitamins

Table 9.1: Process development for antibiotic production in *Streptomyces coelicolor*.

Batch	Medium	pH	Feed mode
A1	A	5.9	Continuous
A2	A	6.9	Continuous
A3	A	5.9	Pulse-pause
A4	A	6.9	Pulse-pause
B1	B	5.9	Continuous
B2	B	6.9	Continuous
C1	C	5.9	Continuous
C2	C	6.9	Continuous

which were only multiplied by a factor of 2. In batches B1-2 the feed was a pure glucose solution. Hence, the concentrations of amino acids, salts and vitamins added at the start of the cultivation should be enough to support both the batch and the fed-batch phases. Therefore the amounts added initially corresponded to the total amount added in A, i.e. the sum of the batch medium and the added feed. In batches C1-2, the batch medium and the feed were the same as in batches B1-2, except for a doubling of the phosphate concentration and a corresponding adjustment of K_2SO_4 and Na_2SO_4 to keep levels of K and Na constant. Furthermore, ${SO_4}^{2-}$ salts were used instead of Cl^- salts to allow in situ sterilization.

9.2.2 Strain and culture conditions

Streptomyces coelicolor A3(2) M145 was used in all the cultivations and was a kind gift from Mervyn Bibb, John Innes Centre, Norwich, UK. The strain was preserved as a spore suspension at -20°C for maximum 6 months. The spore suspension was made by incubating the strain on MS agar plates (Hobbs et al., 1989) for 10 days and harvesting the spores into 20% glycerol (w/w). The bioreactor was inoculated with 1 mL/L spore suspension (10^5 spores mL^{-1}), which was thawed immediately before the inoculation. The cultivations were carried out in stainless steel bioreactors (BIOSTAT Cplus 10-3) (Sartorius, Melsungen, Germany). The initial volume was 5 L which was increased to around 9 L at the end of the fed-batch phase. The bioreactors were operated at 30°C and pH was controlled at either 5.9 or 6.9. The aeration rate was kept approximately constant at 1 vvm throughout the cultivations by adjusting the aeration rate in steps as the volume increased. The agitation was 600 rpm. The feed was started after a sustained drop in the CER was observed. To suppress foam formation, 100 μL/L Antifoam 206 (Sigma-Aldrich, St. Louis, MO, USA) was added at four time points during the cultivations and additionally if formation of foam was observed. The batches were stopped when the feed was depleted. Batches A2 and A4 were terminated earlier, at around 80 h, due to excessive foaming.

9.2.3 Medium

A: The batch medium contained 10 g/L Bacto casamino acids (BD biosciences, MD, USA), 15 g/L glucose, 10 mM KCl, 2mM Na_2SO_4, 4 mM NaH_2PO_4, 2 mM citric acid, 1.25 mM $MgCl_2$, 0.25 mM $CaCl_2$, 5 mL/L trace metals solution, 1 mL/L vitamin solution, and 200 μL/L Antifoam 206 (Sigma-Aldrich, St. Louis, MO, USA). The feed contained the same components as the batch medium all multiplied by a factor of 3 except for trace metals and vitamins which were multiplied by a factor of 2. pH was controlled by addition of 2M NaOH and 1 M H_2SO_4.

B: The batch medium contained 34 g/L Bacto casamino acids (BD biosciences, MD, USA), 10 g/L glucose, 6.5 mM Na_2HPO_4, 7 mM KH_2PO_4, 13 mM $MgSO_4$, 0.85 mM $CaSO_4$, 13 mM K_2SO_4, 4.5 mM Na_2SO_4, 20 mM $(NH_4)_2SO_4$, 7 mM citric acid, 15 mL/L trace metals solution, 6 mL/L vitamin solution, and 200 μL/L Antifoam 206 (Sigma-Aldrich, St. Louis, MO, USA). The feed contained 780 g/L glucose. pH was controlled by addition of 2M NH_4OH and 1 M H_2SO_4.

C: The batch medium contained 34 g/L Bacto casamino acids (BD biosciences, MD, USA), 10 g/L glucose, 10 mM Na_2HPO_4, 16 mM KH_2PO_4, 9 mM $MgSO_4$, 0.85 mM $CaSO_4$, 8 mM K_2SO_4, 20 mM $(NH_4)_2SO_4$, 7 mM citric acid, 15 mL/L trace metals solution, 6 mL/L vitamin solution, and 200 μL/L Antifoam 206 (Sigma-Aldrich, St. Louis, MO, USA). The feed contained 780 g/L glucose. pH was controlled by addition of 2M NH_4OH and 1 M H_2SO_4.

The trace metals solution contained 20 mM $FeCl_3$, 10 mM $CuCl_2$, 50 mM $ZnCl_2$, 10 mM $MnCl_2$, 20 μM Na_2MoO_4, 20 mM $CoCl_2$ and 10 mM H_3BO_4. The vitamin solution contained 50 mg/L biotin, 1 g/L Ca-pantothenate, 1 g/L nicotinic acid, 25 g/L myo-inositol, 1 g/L thiamine-HCl, 1 g/L pyridoxine- HCl and 0.2 g/L para-amino benzoic acid. All medium components except glucose, chloride salts, trace elements and vitamins were autoclaved in the bioreactor. Glucose was autoclaved separately with the chloride salts, and trace elements and vitamins were added to the autoclaved bioreactor through a sterile filter.

9.2.4 Analytical methods

Dry cell weight (DCW) was determined by filtration of a known volume of cultivation broth through a preweighed, predried 0.45 μm pore size filter (PESU membrane, Sartorius Stedim Biotech, Germany). The filter was dried for 15 minutes at 150W in a microwave oven and the weight gain of the filter was determined. For actinorhodin extraction, 1 mL 2 M NaOH was added to 1 mL culture broth, the sample was mixed, and the biomass was separated by centrifugation of the sample at 10000 g for 5 minutes. The actinorhodin concentration was determined by measuring the absorbance of the supernatant at 640 nm, using an extinction coefficient of 25320 $cm^{-1}M^{-1}$. Particle size distribution was measured by the laser diffraction method on a Mastersizer 2000 with a Hydro SM manual small sample dispersion unit (Malvern Instruments Ltd., Worcestershire, UK). The dispersions were made by adding small amounts of cultivation broth to distilled water in the dispersion unit until a laser saturation of approximately 12-15% was obtained. The refractive index of the cell pellets was set to 1.52 and the absrobance to 0.1. Mie theory was applied for deconvolution of the size

distribution. Three successive measurements were made for each sample and the average was calculated. The results were verified by microscopy.

9.2.5 On-line data collection

Feed rate, added acid and base (calculated based on flow rates), weight of feed bottle, temperature, pH and dissolved oxygen concentration were logged in BioPAT MFCS/win (Sartorius, Melsungen, Germany). Carbon dioxide and oxygen in the off-gas were monitored continuously with an acoustic gas analyzer (Innova 1313 Fermentation Monitor, LumaSense Technologies, Denmark) and logged in BioPAT MFCS/win. The classical monitoring data consisted of the variables: time, acid added, base added, feed added, CO_2 concentration in exhaust, O_2 in exhaust, pO_2, feed rate, pH, aeration rate, temperature, volume, OUR, CER, cumulative CER, cumulative OUR, sqrt(cumulative CER), sqrt(cumulative CER), derivative of CER and the cumulative pO_2.

On-line turbidity was measured with a Fundalux II probe (Sartorius, Melsungen, Germany) with an optical path of 10 mm and measurement range between 840-910 nm. The probe was introduced directly into the reactor and connected to the bioreactor control unit. Data were logged in BioPAT MFCS/win.

Fluorescence spectra were collected with a BioView spectrofluorometer (Delta, Hørsholm, Denmark), measuring through a borosilicate glass window placed in the lower part of the bioreactor. Each recorded spectrum was the average of five scans, with excitation wavelengths ranging from 270 nm to 550 nm in steps of 20 nm, and emission wavelengths ranging from 310 nm to 590 nm, also in steps of 20 nm. The data were collected in the BioView control software (Delta, Hørsholm, Denmark). The instrument has been described previously in detail by Marose et al. (1998) and Lindemann et al. (1998).

The dielectric spectroscopic measurements were done with the Biomass Monitor 220 (Aber Instruments, Aberystwyth, UK). The spectrometer was equipped with a 25 mm probe containing four electrodes. The probe was introduced directly into the reactor and sterilized in situ. During the cultivations, 25 frequencies from 0.1 to 20 MHz were scanned every 1 minute and the capacitance as well as the conductance of the cell suspension were registered at each frequency. The data were collected and stored in AberScan (Aber Instruments, Aberystwyth, UK). Due to the significant level of noise, all data were smoothed with respect to time using an exponential moving average given by

$$S_n = (1 - \lambda)Y_n + \lambda S_{n-1} \tag{9.1}$$

where S_n and $S_{n-1}$are the smoothed values at time n and n-1 respectively, Y_n is the measurement at time n, and λ is the smoothing constant. λ was set to 0.97 to provide the best possible smoothing while retaining the dynamics of the signal.

9.2.6 Data analysis

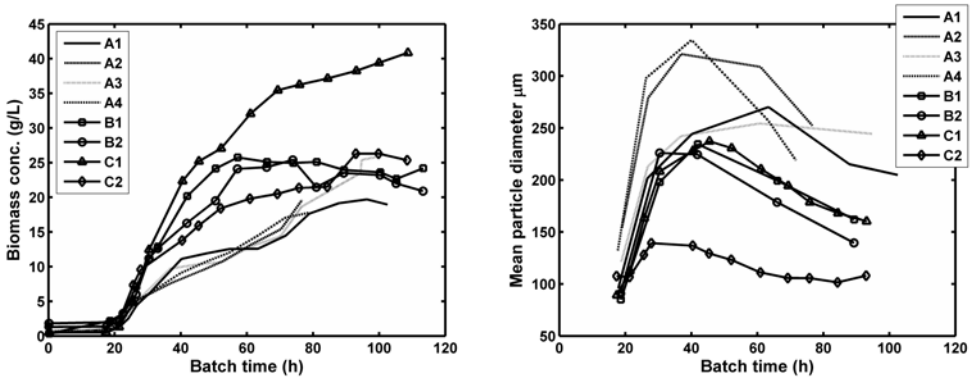
All data analysis was performed in MATLAB (MathWorks Inc, MA, USA) and using the PLS Toolbox for MATLAB (Eigenvector Research Inc., WA, USA). Prior to the analysis, all data sets where checked for outliers by inspection of PCA scores, Hotelling's T^2 and Q-residuals.

Table 9.2: Overview of pretreatments and variable selection methods tested for modeling of multivariable data.

Data	Pretreatment	Variable selection
MWF spectra	-No pretreatment -Subtraction of first spectrum -Mean centering -Autoscaling	-Whole spectrum -Genetic algorithms
Classical monitoring data	-No pretreatment -Autoscaling	-Whole data set -Genetic algorithms
Dielectric spectra	-Subtraction of first spectrum -Subtraction of measurement at high frequency (10085 kHz)	-Whole spectrum -Successive removal of low frequency points

The single variable data (turbidity, cumulative CER and dual frequency DE (measured at 465 kHz and 10085 kHz)) were transformed appropriately and linear prediction models for the biomass concentration were estimated based on a least squares fit. The multivariable data with a low number of variables (combination of sensors) were modeled with standard multiple linear regression models. The multivariable data (dielectric scans, MWF, classical monitoring data) with a high number of variables were modeled using partial least squares (PLS) regression with and without variable selection and various pretreatments of the data. Table 9.2 summarizes the tested data pretreatment and variable selection methods. For the MWF data, the use of subsets of the data for calibration and validation was also evaluated, thus decreasing the variation by only including batches with the same pH or medium. In addition, parallel factor analysis (PARAFAC) was performed of the MWF spectra in order to identify the contributing fluorophores. The theoretical Cole-Cole equation was fitted to the dielectric spectral data (capacitance and conductance) by the method described in Dabros et al. (2009b).

In order to compare the different model types, all model calibration and validation was performed in the same way by using a repeated calibration and validation scheme. The models were calibrated on seven batches and validated on the remaining batch. This was repeated eight times, each time leaving a new batch for validation. The root mean square error of prediction (RMSEP) was calculated for each batch and finally the mean and standard deviation of the RMSEP were calculated based on the results for the eight validation batches.



(a) Dry cell weight **(b)** Mean particle diameter

Figure 9.1: Plot of the development of DCW (a) and mean particle diameter (b) in batches A1-4, B1-2 and C1-2.

9.3 Results

The eight batches attained final biomass concentrations ranging from 18 to 41 g/L DCW (Figure 9.1a). Seven batches reached final biomass concentrations between 18 and 26 g/L DCW whereas the remaining batch (C1) reached 41 g/L DCW and can be considered as an outlier. A very large difference between morphologies of the different batches was observed as indicated by the mean particle diameter of the cell clumps (Figure 9.1b). Furthermore, a change in morphology was observed during each batch. The clump diameter generally increased in the batch phase. Between 30-50 h a population of smaller clumps with diameter around 100-150 μm appeared. This population continued to grow with increasing diameter and covering an increasing percentage of the total biomass volume. Due to the relative increase in the amount of smaller clumps, this appears as an overall decrease in the mean particle diameter (Figure 9.1b). The final antibiotic titer varied from 225 mg/L to 6405 mg/L as a result of the process development work. This illustrates large differences in the secondary metabolism.

9.3.1 Prediction of biomass concentration

Prediction models for the biomass concentration were calibrated based on each of the on-line sensors as well as combinations of the different sensors. A number of different approaches (variable selection, data pretreatment, and partition of the data set) were tested for each of the sensors and the best results were chosen for comparison to the other sensors (Table 9.3). The sensors have different requirements for the calibration data, i.e. the multivariable sensors generally require more data for calibration since more parameters have to be estimated. In order to make a fair comparison between the different sensor types, as a starting point the results are reported based on the use of all available batches. In general, the prediction errors were high for batch C1 and results are also shown without this batch to facilitate comparison. The results of the individual sensors are described

below.

Software sensors

Software sensors were calibrated based on the cumulative CER alone and based on PLS regression of the combined classical monitoring data set. A linear correlation was found between biomass concentration and the square root of the cumulative CER, and a simple linear model was calibrated based on these data. Variable selection using genetic algorithms was tested in combination with PLS regression of the combined data set. A small improvement was found by applying variable selection compared to PLS of the whole data set. However, compared to the simple model based on cumulative CER, no noteworthy improvements were found by including additional process data. The prediction error for the PLS model was 2.2 ± 0.6 g/L compared to 2.0 ± 0.9 g/L for the model based solely on cumulative CER. In fact, when considering the sensors one at a time, the linear regression of the transformed cumulative CER provided the lowest average prediction error of the biomass concentration (Table 9.3). A closer look at the predictions showed that the cumulative CER mainly captured the overall development of the biomass concentration, whereas the dynamics of the growth phases was not so clearly discernible (data not shown). The biomass concentration was generally overestimated towards the end of the batches which may be explained by different biomass/CER yields in this phase.

Turbidity

A non-linear correlation was found between the on-line turbidity measurements and the biomass concentration and a 2^{nd} order polynomial was fitted to the data. The non-linearity may occur close to saturation of the sensor and/or may be influenced by changes in size distribution of the cell pellets, which tended to first increase in the growth phase and then decrease in the production phase (Figure 9.1). Both C1 and C2 showed different correlations between the biomass concentration and the turbidity measurements. For C2 this was clearly caused by a significantly smaller pellet size in this particular batch. For C1 this was most likely caused by non-linear effects close to saturation of the sensor. The on-line turbidity sensor provided prediction errors in the same order of magnitude as DE and the CER (2.3 ± 1.0 g/L compared to 2.8 ± 0.6 g/L and 2.0 ± 0.9 g/L for DE and CER respectively.

Dielectric spectroscopy

The dielectric spectra were very noisy, particularly in the low frequency range, and they were therefore smoothed prior to modeling. It was chosen to use a moving average filter with exponential weights, which is also suitable for on-line estimation. The conductance varied greatly between the different batches. In batches A1-4, the concentration of salts was relatively low initially but increased in the fed-batch phase due to addition of concentrated medium with salts resulting in an increase in conductivity from 7-9 mS/cm to 16-18 mS/cm. Batches B1-2 and C1-2 all started with a high concentration of salts which then decreased during

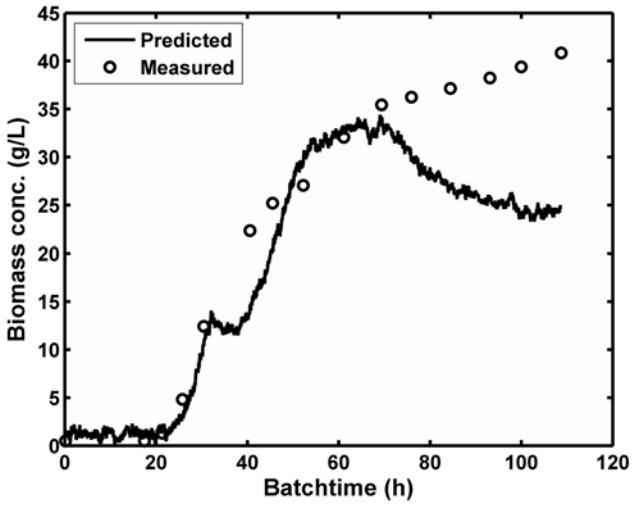


Figure 9.2: Plot of the biomass concentration and the predicted biomass concentration based on dual frequency DE in batch C1.

the cultivation resulting in a decrease in conductivity from 24-27 mS/cm to 8-10 mS/cm. The biomass concentration was estimated based on linear regression and PLS regression of the capacitance data and fitting of the theoretical Cole-Cole equation to the capacitance and conductance data respectively. In all cases the predictions improved when the measurements at 100 to 155 MHz were left out (measurements were subject to noise most likely due to polarization of the probe). The prediction errors for the linear regression and PLS regression were similar in size (2.9 $\pm$ 1.5 for the PLS model compared to 2.8 $\pm$ 0.6 for the linear dual frequency model) and lower than for the Cole-Cole model (data not shown). Both linear and PLS regression provided good predictions in the growth phase capturing the dynamic development of the biomass concentration (Figure 9.2). Even in batch C1, the capacitance sensor correctly predicted the fast growth in the fed-batch phase up to 30 g/L unlike any of the other sensors. However, in most batches the predicted biomass concentration decreased in the production and death phases, although the dry cell weight was stationary or slowly increasing (Figure 9.2). This contributed greatly to the overall prediction errors, which were relatively high (2.8 $\pm$ 0.6 g/L, Table 9.3).

Multi-wavelength fluorescence

A PARAFAC analysis revealed four major peaks (explaining 98% of the variation in the spectra) at the following excitation and emission wavelength combinations: 430/470 nm, 380/450 nm, 430/530 nm and 330/390 nm. The loading profiles of factor 2, 3 and 4 correlate well with the NAD(P)H, flavin nucleotides and pyridoxine respectively (Haack et al., 2004; Ödman et al., 2010) (Food Fluorescence Library, www.models.life.ku.dk). In most batches the intensity of the peaks increased from the beginning of the batch to around 30-45 h, after which the intensity decreased to a stationary level towards the end of the batch. Thus

the signal was not immediately correlated to the biomass concentration. This effect may be the result of a drastic decrease in the concentration of biological fluorophores in the cell mass or/and some quenching effect or inner filter effect of the broth. It was chosen to use PLS regression of the unfolded spectra based on previous results from Ödman et al. (2010). Models were calibrated with mean centering or autoscaling of the spectra prior to the analysis. Furthermore, different partitions of the data set were tested (see materials and methods) along with variable selection using genetic algorithms. The best results were obtained by using all batches together, autoscaling the data and applying genetic algorithms for variable selection. However, the average RMSEP of 4.6 g/L obtained with the MWF sensor was higher than the prediction errors obtained with the other sensors, which ranged between 2-2.8 g/L.

Combination of sensors

Based on the results it was chosen to test a combination of DE and CER as well as DE, CER and the squared turbidity using multiple linear regression. Both combinations of the sensors resulted in a lowering of the prediction error and the combination of the square root of the cumulative CER, DE and the squared turbidity provided the lowest prediction error of all sensors (Table 9.3). The prediction error was thus reduced to 1.5 g/L corresponding to 6% of the covered range of biomass concentrations. This combination was therefore chosen to test the use of an on-line biomass sensor and supervision strategy during process development described below in Section 9.3.2.

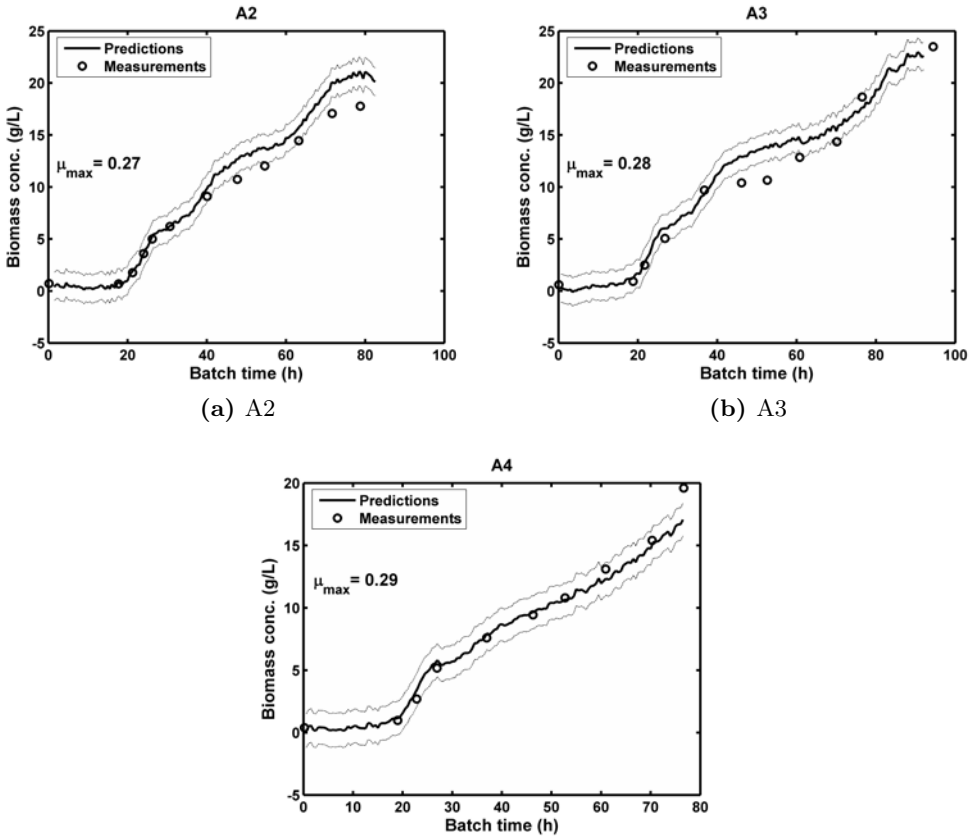
9.3.2 Development and use of on-line biomass sensor during process development

The univariable sensors based on measurements of CER, dual frequency DE and turbidity only require little data for calibration. It was therefore possible to calibrate a biomass sensor based on these signals solely using data from batches A1 and B1. This model was subsequently used to predict the biomass concentration in the remaining batches. The predictions were computed along with a 95 % confidence interval of the predictions. A simple supervision strategy was implemented based on the Hotelling's T^2. The Hotelling's T^2 value was calculated for each measurement point along with the 99% confidence interval based on the calibration data. A value outside the confidence limits indicated that the measurement point was different from the calibration data, and that the sensor estimations should be checked. In addition, the maximum specific growth rate was estimated based on the estimated biomass concentration.

The estimation of the multiple linear regression model resulted in an R^2 of 0.98 which showed that the major part of the variation in biomass concentration in the two calibration batches was captured by the model. The biomass concentration was predicted well in batches A2, A3 and A4 (Figure 9.3). In batches B2 and C2 the Hotelling's T^2 values rose above the 99 % confidence interval, which is indicated by the shaded area (Figure 9.4). The calculation of the Hotelling's T^2 can easily be performed on-line and thus provides the opportunity to adjust the predictions in real-time. A closer look at the individual measurements of

Table 9.3: Comparison of prediction errors for biomass concentration based on the on-line sensors. The average and standard deviations of the RMSEP are calculated based on the repeated model calibration and validation (see Section 9.2.6). ^{A}Without batch C2.

Sensor	Model	mean RMSEP ± std (g/L)	
		All batches Range: 0-41 g/L	Without batch C1 Range: 0-26 g/L
Turbidity	2$^{\text{nd}}$ order polynomial	2.4 ± 1.3^{A}	2.3 ± 1.0^{A}
DE	Linear regression of dual frequency	3.2 ± 1.2	2.8 ± 0.6
DE	PLS	3.4 ± 1.4	2.9 ± 1.5
CER	Linear regression of sqrt(cumulative CER)	3.3 ± 2.4	2.0 ± 0.9
Classical monitoring data	PLS and variable selection with GA	3.4 ± 1.2	2.2 ± 0.5
MWF	PLS	5.9 ± 3.3	4.6 ± 2.4
CER, squared turbidity, DE	Multiple linear regression	1.7 ± 0.6^{A}	1.5 ± 0.4^{A}
CER, DE	Multiple linear regression	2.3 ± 1.6	1.6 ± 0.3

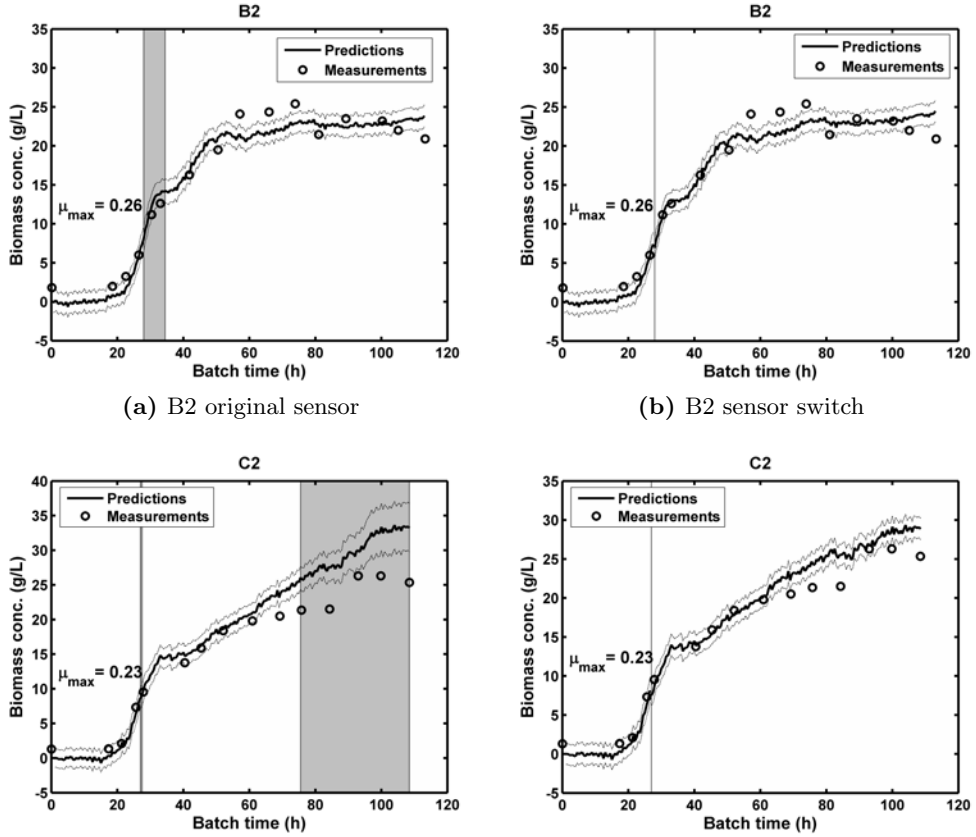


(a) A2 **(b)** A3

(c) A4

Figure 9.3: Predictions of biomass concentration and 95 % confidence intervals for the predictions based on CER, DE and turbidity measurements in batches A2 (a), A3 (b) and A4 (c).

CER, turbidity and DE revealed that the turbidity was larger compared to the measurements of the other two sensors in batches B2 and C2. By the switching to the model based on CER and DE, it was possible to change the predictions (Figure 9.4b and d). In batch B2, the effect of the sensor switch was small, but in C2, the predictions were improved significantly. For C1 the Hotelling's T^2 rose above the confidence interval around t = 50 h (Figure 9.5). A closer look at the individual measurements showed that the capacitance measurements were significantly higher than the two other measurements. However, by experience the capacitance measurements are very accurate during growth and thus it was chosen not to change the model. However, the high Hotelling's T^2 values serve as a warning that the predictions may not be trusted and additional off-line samples can be taken to check the predictions of the model.



(a) B2 original sensor

(b) B2 sensor switch

(c) C2 original sensor

(d) C2 sensor switch

Figure 9.4: Predictions of biomass concentration and 95 % confidence intervals for the predictions based on CER, DE and turbidity measurements for batches B2 and C2. Left: Original predictions based on CER, DE and turbidity. The grey shaded areas indicate the time intervals in which the Hotelling's T^2 values were above the 99 % confidence interval. Right: Predictions resulting from the switch between two different sensors: In the first part predictions are based on CER, DE and turbidity and after the switch, marked by the vertical line, predictions are based on CER and DE.

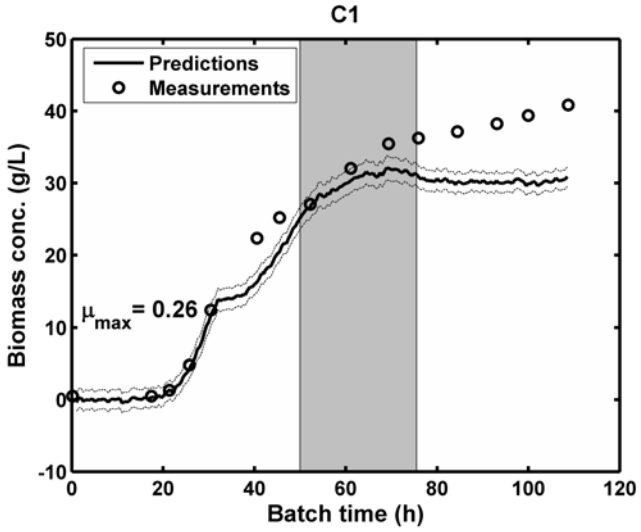


Figure 9.5: Predictions of biomass concentration and 95 % confidence intervals for the predictions based on CER, DE and turbidity measurements in batch C1. The grey shaded areas indicate the periods in which the Hotelling's T^2 was above the 99 % confidence interval.

9.4 Discussion

9.4.1 Multivariable sensors

Comparison of the on-line sensors over a wide range of process conditions showed that the detailed information provided by the advanced sensors, MWF and scanning DE, did not provide better predictions of the DCW and the simpler sensors based on CER, turbidity and dual frequency DE, when applied for estimation of DCW. Using MWF, Ödman et al. (2010) obtained good results for the prediction of biomass concentration in *Streptomyces coelicolor* cultivations over a relatively wide range of process conditions. However, under the increased process variation imposed in our experiments, a characteristic of most process development studies in laboratories or pilot plants, MWF failed to produce reliable estimations of DCW. This may be explained by the differences in pH, which are known to have a large effect on the fluorescent properties of fluorophores (Valeur, 2001). Moreover, cell metabolism was highly varied in the batches as observed from the growth curves and the antibiotic production. Thus it is likely that the correlation between DCW and the biological fluorophores such as NAD(P)H and flavins was not the same in all batches.

For DE, Dabros et al. (2009b) found it advantageous to use the full spectrum for prediction of biomass concentration under noisy conditions, which were caused by interference from reactor components in small-scale reactors. We did not come to a similar conclusion, and the explanation for this may be either the substantial smoothing applied prior to modeling in our work or lower noise levels. Overall it is likely that the more advanced sensors will perform better under less varying process conditions, where they have the potential to provide additional

information, e.g. about cell metabolism or cell morphology (Dabros et al., 2009b; Hantelmann et al., 2006; Siano, 1997). The more advanced sensors could, for example, be implemented during the later stages of process development where only smaller changes are made to the protocol.

9.4.2 Univariable sensors

Predictions of the biomass concentration based on the on-line turbidity measurements are influenced by the morphology of the cells since the scattering properties are related to particle size and shape (Willams and Norris, 1987). Nonetheless, the predictions of the biomass concentration were relatively accurate over a wide range of morphologies. In fact, large prediction errors mainly occurred in batch C2, in which the morphology of the cells was significantly different from the other batches (Figure 9.1b).

Dual frequency DE provided accurate predictions of the DCW during growth, but generally gave relatively large prediction errors in the production and death phases. This is in agreement with previous observations in filamentous systems (Neves et al., 2000; Sarra et al., 1996) and in cell cultures, where the effect was partly attributed to decreases in cell viability, decreases in cell size, and changes in the capacitance-per-unit membrane area (Ansorge et al., 2007; Noll and Biselli, 1998). Alternatively, the effect might have been caused by changes in the ratio between cell mass and cell volume, since DE measurements are proportional to the cell volume (Harris et al., 1987).

The prediction of biomass concentration based on the CER is an indirect method, which is based on the assumption that the biomass/CER yield is constant. However, the predictions of biomass concentration based on CER were robust over a relatively wide range of process variation. In fact the sensor based on CER performed as good or better than the more direct sensors over the range covered by A1-4, B1-2 and C2.

Although the capacitance and the turbidity sensors directly measure a physical property of the biomass, the prediction of DCW is still based on correlations which are subject to changes caused by differences in morphology, cell viability, etc. It can be discussed whether the DCW is the best measure of biomass concentration since it does not differentiate between viable or dead cell mass. In some applications it may be that the information provided from the on-line sensors is actually a better measure of biomass, for example dielectric measurements of viable biomass (Guan et al., 1998; Neves et al., 2000). However, in this study, it was important to use one established measure of biomass concentration as a reference to facilitate the comparison. When the advanced biomass sensors have won wider acceptance in the industry, it is likely that the sensor measurements will be used directly as a measure of biomass concentration.

9.4.3 Combination of sensors

The CER, DE and turbidity can all be used individually to provide relatively accurate estimations of DCW over a wide range of process conditions. However, the predictions were improved by combining the three different sensors measuring different properties of the biomass, thus creating a positive averaging effect.

Furthermore, differences between the sensor signals may provide additional information. A discrepancy between the predictions based on capacitance and CER on the one hand, and the turbidity probe on the other hand, suggests a different morphology, while differences between predictions based on CER and the other signals suggest changing yields. Finally, the use of multiple sensors makes it possible to check the individual predictions and make adjustments *on-line,* as was demonstrated in Section 9.3.2. A combination of sensors has previously been used for the estimation of biomass concentration in cultivations. Using advanced data reconciliation, Dabros et al. (2009a) combined DE, off-gas analysis, and mid infrared spectroscopy for the prediction of biomass and metabolite concentrations in *Saccharomyces cerevisiae* cultivations under more defined process conditions. Our results have shown that a simple combination of sensors is advantageous under more varying process conditions.

9.4.4 Application in process development

Process development represents a very important area for the application of on-line sensors in the industry. The on-line sensors can provide information about the process dynamics which may be of high value in process optimization. Furthermore, the process variation will generate valuable experience and help to establish the range of validity of the sensors. However, the high level of variation requires very robust sensors. We have demonstrated that a biomass sensor based on measurements of CER, dual frequency DE and turbidity can be calibrated and subsequently used to provide reliable biomass measurements *during* a short process development. Its usability was further improved by combination with estimated confidence intervals of the predictions and calculated Hotelling's T^2 values, which provided information about the reliability of the predictions. This allowed the supervision of the sensor and presented the option of switching between individual sensors in real-time. Biomass measurements can, for example, be used to estimate the specific growth rate on-line, as it was also demonstrated.

9.5 Conclusion

The aim of this study was to compare some of the most prominent on-line biomass sensors during the development of a filamentous cultivation process. The results showed that the more robust univariable sensors (based on measurements of CER, turbidity, or dual frequency DE) provided better or no worse predictions of DCW than the more advanced multivariable sensors (based on MWF, scanning DE, multivariable modeling of classical monitoring data). Furthermore, it was shown that a combination of the sensors reduced the average and the standard deviation of the prediction errors (2.0 ± 0.9 g/L to 1.5 ± 0.4 g/L) and provided a means of checking the sensor in real-time, thus allowing operator intervention. Finally, it was demonstrated that the biomass sensor could be calibrated and subsequently used to provide reliable biomass measurements during a short process development. This study has therefore increased our knowledge of on-line biomass sensors for more challenging conditions, such as those observed during early process development. The process development stage represents a large

potential for future applications of on-line sensors in industry.

Acknowledgements

Acknowledgement is made of Insatech A/S and Aber Instruments Ltd. for the use of the Biomass Monitor 220. The Ph.D. project of Nanna Petersen Rønnest is supported by a grant from the Innovative Bioprocess Technology Research Consortium financed by the Danish Research Council for Technology and Production Sciences, Chr. Hansen A/S, Danisco A/S and Novozymes A/S.

Bibliography

Ansorge S, Esteban G, Schmid G. 2007. On-line monitoring of infected Sf-9 insect cell cultures by scanning permittivity measurements and comparison with off-line biovolume measurements. Cytotechnology 55:115–24.

Arnoux AS, Preziosi-Belloy L, Esteban G, Teissier P, Ghommidh C. 2005. Lactic acid bacteria biomass monitoring in highly conductive media by permittivity measurements. Biotechnology Letters 27:1551–1557.

Clementschitsch F, Kern J, Pötschacher F, Bayer K. 2005. Sensor combination and chemometric modelling for improved process monitoring in recombinant *E. coli* fed-batch cultivations. Journal of Biotechnology 120:183–197.

Dabros M, Amrhein M, Bonvin D, Marison IW, von Stockar U. 2009a. Data reconciliation of concentration estimates from mid-infrared and dielectric spectral measurements for improved on-line monitoring of bioprocesses. Biotechnology Progress 25:578–588.

Dabros M, Dennewald D, Currie DJ, Lee MH, Todd RW, Marison IW, von Stockar U. 2009b. Cole-Cole, linear and multivariate modeling of capacitance data for on-line monitoring of biomass. Bioprocess and Biosystems Engineering 32:161–173.

Guan Y, Evans PM, Kemp RB. 1998. Specific heat flow rate: an online monitor and potential control variable of specific metabolic rate in animal cell culture that combines microcalorimetry with dielectric spectroscopy. Biotechnology and Bioengineering 58:464–477.

Haack MB, Eliasson Lantz A, Mortensen PP, Olsson L. 2007. Chemometric analysis of in-line multi-wavelength fluorescence measurements obtained during cultivations with a lipase producing *Aspergillus oryzae* strain. Biotechnology and Bioengineering 96:904–913.

Haack MB, Eliasson Lantz A, Olsson L. 2004. On-line cell mass monitoring of *Saccharomyces cerevisiae* cultivations by multi-wavelength fluorescence. Journal of Biotechnology 114:199–208.

Hantelmann K, Kollecker M, Hüll D, Hitzmann B, Scheper T. 2006. Two-dimensional fluorescence spectroscopy: A novel approach for controlling fed-batch cultivations. Journal of Biotechnology 121:410–417.

Harris CM, Todd RW, Bungard SJ, Lovitt RW, Morris JG, Kell DB. 1987. Dielectric permittivity of microbial suspensions at radio frequencies: a novel method for the real-time estimation of microbial biomass. Enzyme and Microbial Technology 9:181–186.

Hobbs G, Frazer CM, Gardner DCJ, Cullum JA, Oliver SG. 1989. Dispersed growth of *Streptomyces* in liquid culture. Applied Microbial Biotechnology 31:272–277.

Jenzsch M, Gnoth S, Kleinschmidt M, Simutis R, Lübbert A. 2007. Improving the batch-to-batch reproducibility of microbial cultures during recombinant protein production by regulation of the total carbon dioxide production. Journal of Biotechnology 128:858–867.

Jenzsch M, Simutis R, Eisbrenner G, Stueckrath I, Lübbert A. 2006. Estimation of biomass concentrations in fermentation processes for recombinant protein production. Bioprocess and Biosystems Engineering 29:19–27.

Kiviharju K, Salonen K, Moilanen U, Eerikaeinen T. 2008. Biomass measurement online: the performance of in situ measurements and software sensors. Journal of Industrial Microbiology and Biotechnology 35:657–665.

Kiviharju K, Salonen K, Moilanen U, Meskanen E, Leisola M, Eerikaeinen T. 2007. On-line biomass measurements in bioreactor cultivations: comparison study of two on-line probes. Journal of Industrial Microbiology and Biotechnology 34:561–566.

Lantz AE, Jørgensen P, Poulsen E, Lindemann C, Olsson L. 2006. Determination of cell mass and polymyxin using multi-wavelength fluorescence. Journal of Biotechnology 121:544–554.

Li JK, Asali EC, Humphrey AE, Horvath JJ. 1991. Monitoring cell concentration and activity by multiple excitation fluorometry. Biotechnology progress 7:21–27.

Lindemann C, Marose S, Nielsen HO, Scheper T. 1998. 2-dimensional fluorescence spectroscopy for on-line bioprocess monitoring. Sensors and Actuators B: Chemical 51:273–277.

Madrid RE, Felice CJ. 2005. Microbial biomass estimation. Critical Reviews in Biotechnology 25:97–112.

Markx GH, Davey CL, Kell DB. 1991a. The permittistat a novel type of turbidostat. Journal of General Microbiology 137:737–744.

Markx GH, Davey CL, Kell DB, Morris P. 1991b. The dielectric permittivity at radio frequencies and the bruggeman probe: novel techniques for the on-line determination of biomass concentrations in plant cell cultures. Journal of Biotechnology 20:279–90.

Marose S, Lindemann C, Scheper T. 1998. Two-dimensional fluorescence spectroscopy: a new tool for on-line bioprocess monitoring. Biotechnology Progress 14:63–74.

Neves AA, Pereira DA, Vieira LM, Menezes JC. 2000. Real time monitoring biomass concentration in *Streptomyces clavuligerus* cultivations with industrial media using a capacitance probe. Journal of Biotechnology 84:45–52.

Noll T, Biselli M. 1998. Dielectric spectroscopy in the cultivation of suspended and immobilized hybridoma cells. Journal of Biotechnology 63:187–198.

November EJ, Van Impe JF. 2000. Evaluation of on-line viable biomass measurements during fermentations of *Candida utilis*. Bioprocess Engineering 23:473–477.

Ödman P, Johansen C, Olsson L, Gernaey KV, Lantz A. 2010. Sensor combination and chemometric variable selection for online monitoring of *Streptomyces coelicolor* fed-batch cultivations. Applied Microbiology and Biotechnology 86:1745–1759.

Ödman P, Johansen CL, Olsson L, Gernaey KV, Lantz AE. 2009. On-line estimation of biomass, glucose and ethanol in *Saccharomyces cerevisiae* cultivations using in-situ multi-wavelength fluorescence and software sensors. Journal of Biotechnology 144:102–112.

Olsson L, Nielsen J. 1997. Online and in situ monitoring of biomass in submerged cultivations. Trends in Biotechnology 15:517–522.

Sarra M, Ison AP, Lilly MD. 1996. The relationships between biomass concentration, determined by a capacitance-based probe, rheology and morphology of *Saccharopolyspora erythraea* cultures. Journal of Biotechnology 51:157–165.

Siano SA. 1997. Biomass measurement by inductive permittivity. Biotechnology and Bioengineering 55:289–304.

Sundström H, Enfors SO. 2008. Software sensors for fermentation processes. Bioprocess and Biosystems Engineering 31:145–152.

Valeur B. 2001. Molecular fluorescence. Weinheim: Wiley-VCH Verlag.

Willams P, Norris K, editors. 1987. Near-Infrared Technology in the Agricultural and Food Industries. St. Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Xiong ZQ, Guo MJ, Guo YX, Chu J, Zhuang YP, Zhang SL. 2008. Real-time viable-cell mass monitoring in high-cell-density fed-batch glutathione fermentation by *Saccharomyces cerevisiae* T65 in industrial complex medium. Journal of Bioscience and Bioengineering 105:409–413.

Chapter 10

Guidelines for industrial application of advanced sensors

10.1 Introduction

The purpose of this chapter is to provide some guidelines for future applications of on-line sensors in industrial cultivation processes based on the case studies described in the previous chapters and relevant studies in the literature. The development of an on-line sensor application requires careful consideration of a number of aspects. Firstly, it is important to consider the requirements for the sensor. This includes identification of key process parameters and considerations of the process conditions, which influences measurements and calibration of the sensor. The measurements are influenced by process characteristics such as the type of microorganism used, the aeration, stirring etc. as well as the level of variation in the process. This is highly dependent on the application area, thus pilot plant or laboratory process development has different requirements than large scale production. After formulation of the requirements, a suitable sensor can be chosen. Most advanced on-line sensors necessitate careful calibration and validation. It is important consider the experimental design in order to obtain a calibration which fulfills the requirements for analyte-specificity and robustness. Finally, the sensor can be implemented - possibly in combination with a supervision system to compensate for uncertainty in the predictions and/or drifts in the process or sensor. In the following, the different aspects of the development of on-line sensor applications are discussed step-by-step. This is not a complete presentation of all the aspects of on-line sensor calibration. Instead the discussion focuses on the aspects which have come up as particularly important in the research presented in this thesis.

10.2 Identification of key process parameters

Development of monitoring schemes with advanced sensors is laborious and time consuming. The first question that should be answered is therefore: what can be gained by on-line monitoring? And what should be monitored in order to obtain improvements in the process operation? These decisions should be made based on knowledge of the process. In reality, little is often known about the details of the process. In that case the development of on-line monitoring applications can be part of an iterative procedure where the on-line sensors provide new information about the process, which is then used to identify an improved monitoring scheme. Such an iterative procedure could for instance be integrated in the development of new processes/products. The common monitoring schemes applied for bioprocess

monitoring can be divided in two groups: (1) Monitoring of specific analytes based on correlation models calibrated for the on-line sensors and (2) unsupervised monitoring based on the establishment of normal operation trajectories of the sensor signals.

10.2.1 Monitoring of specific analytes

Advanced on-line sensors are often used for the measurement of important analytes such as substrates, products or biomass. Measurements of nutrient concentrations can be used to control the nutrient levels within an optimal range (González-Vara et al., 2000; Tosi et al., 2003). Measurements of the product concentration can be used to calculate on-line estimates of production rate, optimize conditions for production and determine when the process should be terminated. On-line estimates of the biomass concentration can be used to control the biomass concentration at a constant level (Markx et al., 1991) or can for example be used to calculate specific production rates and help to identify optimal growth and production conditions (Navrátil et al., 2005). The success of the on-line monitoring depends on the identification of important process parameters and the availability of a suitable monitoring strategy. If detailed knowledge of the process is available, different tools exist, such as ICAS-PAT (Singh et al., 2009, 2010), for the identification of the limiting parameters.

In many cases, detailed knowledge of the process is not available prior to the start up of the monitoring project. In that case, it can be considered to start with an at-line application. The results presented in this thesis have shown that for NIR spectroscopy, at-line measurements of the filtered cultivation broth were useful to predict analyte concentrations such as glucose and ammonium, and this with a higher accuracy than the on-line applications (Chapter 8). By starting with an at-line application, it will be possible to obtain important information of nutrient and product concentration in almost real-time. The analysis is very fast and the sample can be analyzed the minute it is taken from the bioreactor, which will decrease influence from storage and shipment to laboratory facilities. This will generate knowledge of the process and help to establish the important process parameters and at the same time provide an evaluation of the applicability and usefulness of the sensor technology. The at-line application is cost efficient since it can serve a large number of bioreactors at the same time. Furthermore, it is much easier and faster to perform designed experiments for the sensor calibration (Riley et al., 1998b). Finally, by allowing some sample pretreatment it is possible to remove the biomass from the samples making it easier to calibrate a model which covers several cultivation processes. If necessary, the at-line application can be transferred to on-line, although this is far from straightforward as it was illustrated in Chapter 8. Thus the performance of the at-line application will most likely serve as a 'best case' scenario.

10.2.2 Unsupervised monitoring

The signals from the on-line sensors do not need to be correlated to known analytes but can be used directly for monitoring of the process. For example, process trajectories can be established based on the sensor signals and used to identify

phase shifts or process faults such as contamination of the cultivation (Haack et al., 2007; Navrátil et al., 2005). One of the main advantages of unsupervised monitoring is that it does not require analysis of a large number of off-line samples, a task that is otherwise known to be quite laborious. In addition, it is possible to utilize all of the information in the sensor signals without discarding important information because it is not correlated to any of the measured analytes. The main disadvantage of unsupervised monitoring is that the observed changes or differences cannot be directly linked to changes in for example analyte concentrations and thus may not contribute to an increased mechanistic understanding of the process. However, unsupervised monitoring may still provide valuable process knowledge, such as determination of correlations between process behavior and measured process parameters and conditions.

In order to obtain useful results it is critical, particularly for sensors such as near infrared spectroscopy (NIR), to identify a suitable pretreatment of the data and/or to select the most important latent components, which extracts important process information. In our research, the first two components of PCA of the NIR spectra were mainly correlated to scattering properties of the broth thus primarily reflecting changes in biomass. Small differences in metabolite concentrations may thus be difficult to discern in the NIR spectra.

Some of the advanced on-line sensors are particularly suitable for unsupervised monitoring schemes. Classical monitoring data, which are already collected and thus do not constitute a major capital investment, can be of great use in combination with multivariable statistical process control (MSPC) (Albert and Kinley, 2001). Multi-wavelength fluorescence spectroscopy (MWF) is another example of a sensor technology which may be of great use in unsupervised monitoring schemes. In Chapter 9 it was not possible to establish a correlation between the MWF signal and the dry cell weight due to the large variations in the process. Instead, the sensor can be used in unsupervised monitoring schemes to provide information about changes in cell metabolism (Hantelmann et al., 2006). The MWF data are particularly suitable for unsupervised modeling, because the three-way structure allows the use of tools such as parallel factor analysis (PARAFAC) for identification of the "true" underlying chemical composition, and thus eases interpretation of the spectra. Finally, NIR and MIR spectroscopy, can be useful in unsupervised monitoring applications, since the spectra contain information about most of the substrates, metabolites and products in the cultivation broth.

10.3 Selection of sensor

It is difficult to establish general guidelines for the selection of suitable sensors for cultivation processes because the performance of the individual sensors is highly dependent on the specific system. This section presents a brief overview of some of the available technologies, the range of application, advantages and disadvantages to provide some guidance for the selection. For a more detailed description readers are referred to review articles such as Vaidyanathan et al. (1999), Schügerl (2001) and Clementschitsch and Bayer (2006).

10.3.1 Measurement of substrates, metabolite and product concentration

For in situ measurements of substrate, metabolite and product concentrations above, vibrational spectroscopic methods such as NIR and mid infrared spectroscopy (MIR) are very suitable. The vibrational spectroscopic measurements are non-destructive, rapid and can supply multi-analyte information. Table 10.1 summarizes the prediction errors obtained in situ for various important analytes using NIR and MIR. Only studies in which the models have been validated on an independent test set are included. The prediction errors originate from different studies and thus cannot be compared directly. However, the relative size of the prediction errors gives some indication of the expected accuracy of the different sensors. The advantage of MIR over NIR is a higher sensitivity and better resolution of the spectra allowing detection of analytes in lower concentrations, discrimination between analytes of similar structure and detection and identification of unexpected compounds (Roychoudhury et al., 2006b; Schenk et al., 2008). Furthermore, the better resolved MIR spectra facilitate the use of synthetic samples and library based methods for calibration, which decreases the number of experiments required for calibration significantly (Schenk et al., 2008). However, MIR has some disadvantages concerning practical implementation. First of all, light in the MIR range is transported very poorly in optical fibers, which limits the distance between detector and sample to 1-5 m, and which makes multiplexing of probes difficult. Secondly, attenuated total reflectance (ATR) which is the measurement principle applied for in situ monitoring applications, uses a very short path length for measuring the sample (2-5 μm). This makes it very sensitive to fouling (Anders Larsen, Q-Interline, Personal communication). Finally, the high water sensitivity may cause some problems for in situ measurements due to the dilute nature of the cultivation broth.

Raman spectroscopy is another vibrational spectroscopic method which has been used for in situ monitoring of cultivation processes. The main advantages of Raman spectroscopy are the low interference from water, flexibility in the choice of wavelengths and the availability of fiber optic probes (Lee et al., 2004). However, a significant disadvantage of Raman spectroscopy is that there is significant interference from fluorescent compounds in the Raman spectra and a reported low signal to noise ratio (Sivakesava et al., 2001). There are only few studies reporting the results of Raman for monitoring of cultivation processes. Further development of robust probes and instruments is needed and it seems that NIR and MIR are superior in most applications at the present time (Clarke et al., 2005; Lee et al., 2004; Sivakesava et al., 2001).

Finally, there are cases where the specific properties of the analytes allows the utilization of more specific sensors. This includes fluorescent compounds (MWF), volatile compounds (electronic nose, gas chromatography, mass spectrometry) or colored compounds (UV/visible spectroscopy) (Table 10.2).

Table 10.1: Overview of prediction errors for different common analytes reported in the literature for NIR and MIR spectroscopy. [A]The systems are roughly divided into simple and complex systems, where the complex systems are cultivations characterized by high(er) biomass concentrations and intense aeration and stirring.

Analyte	System[A]	Prediction error (concentration range)	
		NIR	MIR
Glucose	Simple	0.26 g/L (0-4.12 g/L) (Navrátil et al., 2005)	0.1-0.18 g/L (0-3.6 g/L) Off-set corrected (Rhiel et al., 2002c)
	Complex	2.0 g/L (0-45 g/L) (Petersen et al., 2010)	0.86-4.92 g/L (0-65 g/L) (Dabros et al., 2007; Schenk et al., 2007; Veale et al., 2007)
Glycerol	Complex	3.03 g/L (0-35 g/L) (Holm-Nielsen et al., 2008)	0.6-1.13 g/L (0-13 g/L) (Dabros et al., 2007; Schenk et al., 2008)
Acetate	Simple	0.28 g/L (0-3.77 g/L) (Navrátil et al., 2005)	-
	Complex	0.5-0.6 g/L (0-19 g/L) (Tosi et al., 2003)	0.1-0.33 g/L (0-6 g/L) (Kornmann et al., 2004; Schenk et al., 2008)
Lactate	Simple	-	0.08-0.30 g/L (0-3.6 g/L) (Rhiel et al., 2002c)
	Complex	1.4-1.64 g/L (0-24 g/L) (Tosi et al., 2003)	-
Ethanol	Complex	0.42% (0-7.5%) (Cavinato et al., 1990)	0.92-2.45 (0-30 g/L) (Dabros et al., 2007; Schenk et al., 2007)
Ammonium	Complex	0.20 g/L (0.72-1.98 g/L) (Petersen et al., 2010)	0.1-0.38 g/L (0-2.16 g/L) (Dabros et al., 2007; Kornmann et al., 2004; Schenk et al., 2007, 2008)
Phosphate	Complex	-	0.24-1.14 g/L (2-5 g/L) (Dabros et al., 2007; Kornmann et al., 2004)

Table 10.2: Overview of existing sensor technologies for bioprocess monitoring.

Sensor	Analytes	Process connection	Advantages	Disadvantages
Near infrared spectroscopy	Substrates, metabolites, products and biomass.	In situ and at-line	Measurements are rapid, non-destructive, non-intrusive, provide multianalyte information and require little or no sample preparation (Sivakesava et al., 2001).	Interference from water, broad and overlapping absorption peaks (Sivakesava et al., 2001)
Mid infrared spectroscopy	Substrates, metabolites, products and biomass.	In situ and at-line	Measurements are rapid, non-destructive, non-intrusive, provide multianalyte information, high sensitivity and resolution, require little or no sample preparation (Roychoudhury et al., 2006a; Sivakesava et al., 2001).	Interference from water and signal attenuation in optical fibers (Roychoudhury et al., 2006a).
Raman spectroscopy	Substrates, metabolites and products.	In situ and at-line	Measurements are rapid, non-destructive, non-intrusive and require little or no sample preparation. Low interference from water (Lee et al., 2004)	Interference from fluorescent compounds and low signal to noise ratio (Sivakesava et al., 2001).
Multi-wavelength fluorescence spectroscopy	Fluorescent compounds such as NAD(P)H, ATP, tryptophan and flavins, related for examples to biomass and protein products	In situ and at-line	Measurements are very sensitive, selective, rapid, non-destructive, and require no sample preparation.	Sensitive towards changes in pH, temperature, viscosity and amount of scattering particles.

Continued on next page

Sensor	Analytes	Process connection	Advantages	Disadvantages
Dielectric spectroscopy	Biomass	In situ and at-line	Selective towards live biomass, wide linear range, insensitive to other particulate matter	Sensitive to changes in medium conductance and changes in dielectric properties of the cells.
Flow injection analysis	Dependent on the applied sensor. For example low molecular medium components, enzyme activity or biopolymers	On-line ex-situ through membrane sampling.	High accuracy, short response time, high flexibility and inexpensive components. (Schügerl, 2001; Vaidyanathan et al., 1999)	Require sampling, lack of reliable supervisory systems for on-line fault detection and correction (Schügerl, 2001).
Nuclear magnetic resonance	A wide range of compounds such as amino acids, organic acids, sugars and alcohols.	At-line	Relatively fast analysis time, possible to measure in vivo for metabolic flux analysis (Correia et al., 2005; Schügerl, 2001)	Difficult to apply on-line for bioreactor monitoring (Vaidyanathan et al., 1999)
Ion and liquid chromatography	A wide range of compounds such as amino acids, organic acids, sugars and alcohols.	At-line and off-line. Samples taken from the bioreactor usually pretreated by removal of biomass and precipitation of proteins. Few examples of on-line applications (Buttler et al., 1994, 1996; Torto et al., 1997).	High precision, flexibility and accuracy (Vaidyanathan et al., 1999)	Long analysis time, requires sample pretreatment: biomass removal and protein precipitation (Schügerl, 2001).

Continued on next page

Sensor	Analytes	Process connection	Advantages	Disadvantages
Electronic nose	Volatile compounds, such as ethanol and complex hydrolysates, and microorganisms such as *E. coli* and yeasts.	On-line ex-situ. Analysis of off-gas	Suitable for detection of contamination and prediction of metabolic state. Information can be correlated to non-volatile and low concentration compounds (Liden et al., 2000).	Selective towards volatile compounds, further development require sensors with higher stability, less signal drift, greater control of irreversible effects and shorter recovery time than those available at present (Bachinger et al., 2000)
Mass spectrometry	Volatile and low molecular weight compounds.	On-line ex-situ. Analysis of off-gas or though sampling from bioreactor.	Easy automation and low detection limit. (Clementschitsch and Bayer, 2006; Schügerl, 2001)	Require sampling from the bioreactor for determination of concentrations in broth.
Gas chromatography	Gaseous and volatile compounds	Few examples of on-line ex-situ. Analysis of exhaust gas or through membrane sampling.	Simultaneous measurements of multiple analytes.	GC is more difficult to automate compared to mass-spectrometry (Schügerl, 2001).
Microcaliometry	Heat flow rate which is a reflection of metabolic rate.	On-line ex-situ. Samples analyzed in by-pass loop.	Provides global measure of cellular activity (Vaidyanathan et al., 1999).	Low signal resolution, instruments are relatively expensive. (Guan et al., 1998; Schügerl, 2001; Vaidyanathan et al., 1999)

Continued on next page

Sensor	Analytes	Process connection	Advantages	Disadvantages
On-line microscopy	Biomass and morphology	in situ or on-line	Provide detailed information of cell morphology.	Reliant on advanced image analysis tools for rapid quantitative analysis of the images in order to distinguish the biomass from air bubbles and other particulate matter present in the broth. Sensitive equipment.

There are large differences between the errors reported for the simple systems (characterized by low biomass concentration, gentle stirring and aeration, and clear medium) and the complex cultivation systems (characterized by higher biomass concentrations and intense stirring and aeration) (Table 10.1). In situ sensor technologies are constantly improving; however, it may be difficult to achieve the desired sensitivity and accuracy with the in situ technologies. Before investing considerable time and money, it should be carefully considered if it is likely that an in situ sensor can provide measurements of sufficiently high quality. For example, limiting concentrations of substrates and concentrations of biproducts are often in the < 1 g/L range (Nielsen et al., 2003). If a high accuracy is necessary, it can be considered to use flow injection analysis (FIA) combined with a suitable sensor. FIA provides accurate and flexible measurements with a short response time and is relatively cheap (Schügerl, 2001; Vaidyanathan et al., 1999). On-line application of FIA requires automatic sampling from the bioreactor. Various automatic sampling units exist, which allows for some sample pretreatment, for example based on filtration and dialysis (Lapa et al., 2003; Poulsen et al., 2004). However, sampling from the bioreactor is connected to some practical challenges and include a risk of compromising the sterile environment (Vaidyanathan et al., 1999). Furthermore, the technology suffers from a lack of reliable supervisory systems for on-line fault detection and correction (Schügerl, 2001). Alternatively, it can be considered to use at-line analysis, which increases the number of available technologies. Furthermore, at-line analysis facilitates sample pretreatment, e.g. removal of biomass, which can improve the accuracy of the predictions significantly (Chapter 8, Petersen et al. (2010); Schenk et al. (2008)). Available technologies for at-line analysis of substrate, metabolite and product concentrations include the vibrational spectroscopic methods, nuclear magnetic resonance, flow injection analysis and mass spectrometry (Table 10.2).

10.3.2 Biomass measurements

On-line biomass measurements have received much attention over the past years but remain challenging due to the complex and non-constant nature of cellular material. The challenges of biomass measurement have been reviewed by Olsson and Nielsen (1997), Kiviharju et al. (2008) and Madrid and Felice (2005). Methods used for on-line measurement of biomass concentration include dielectric spectroscopy, multi-wavelength fluorescence, on-line turbidity as well as the vibrational spectroscopic methods (Table 10.3).

Dielectric spectroscopy is specifically designed to measure live biovolume and excellent results have been obtained for measurement of biomass concentration in both unicellular and filamentous systems (Ansorge et al., 2007; Dabros et al., 2009b; Xiong et al., 2008). In addition, scanning dielectric spectroscopy can provide information concerning cell size and distribution. Dielectric spectroscopy covers a wide concentration range and has the advantage of being selective towards live cells. The disadvantages include sensitivity to changes in medium conductance and changes in the dielectric properties of the cells.

Multi-wavelength fluorescence measures only fluorescent compounds and is thus able to measure a lower number of compounds but with a very high sensitivity. Measurements of fluorophores such as NADH, tryptophan and riboflavins, which

are present in cellular material, have enabled monitoring of biomass concentration using MWF spectroscopy. The prediction of biomass concentration using MWF spectroscopy is based on a constant or at least predictable correlation between the fluorophores and the biomass concentration, which may only exist under a limited range of process conditions (Haack et al., 2007; Lantz et al., 2006; Ödman et al., 2009). However, the main advantage of MWF is perhaps not the ability to measure biomass concentration but rather the detailed information about the biological fluorophores, which can provide information concerning cell metabolism, such as the redox state of the cells, which is not available with other techniques (Hantelmann et al., 2006).

On-line turbidity measurements are based on the scattering of light by suspended particles and constitute a standard method for determination of biomass concentration off-line. A range of commercial probes are available. The sensors are best suited for measurement of unicellular biomass, for which the sensors provide robust and accurate measurements (Kiviharju et al., 2007). However, the sensors are subject to saturation, which limits the range of biomass concentrations covered by the sensor (Kiviharju et al., 2008; Vaidyanathan et al., 1999).

Although not specific for biomass measurements, the vibrational spectroscopic methods have also been used for estimation of biomass concentration. The measurements can be based on the scattering of light by the biomass particles (Ge et al., 1994) or specific absorption bands for biomass (Vaidyanathan et al., 2001). One of the main advantages of using spectroscopic methods such as NIR, MIR or MWF for biomass estimation, is that they can provide measurements of other important analytes simultaneously.

Less frequently used sensors for measurement of biomass and biomass-related parameters include the electronic nose, on-line microscopy and microcalorimetry. The electronic nose measures the concentration of various volatile compounds and can be applied on-line for measurements of the off-gas. It has been used for classification and prediction of biomass concentration of microorganisms such as yeast and *E. coli* (Gibson et al., 1997; Mandenius et al., 1997). It has proved to be very useful for detection of contaminations and phase shifts during cultivation (Bachinger et al., 2000; Namdev et al., 1998). Microscopy has been developed for on-line and in situ use (Suhr et al., 1995; Treskatis et al., 1997). On-line microscopy offers a detailed characterization of cellular morphology, which can be linked to rheological properties of the cultivation broth as well as the cell metabolism (Riley et al., 2000). Utilization of this technology relies on advanced image analysis tools for rapid quantitative analysis of the images in order to distinguish the biomass from air bubbles and other particulate matter present in the broth. Furthermore, application in an industrial environment sets high requirements for the robustness, which may be a problem for some of the sensitive microscopes. Microcalorimetry measures the heat flow rate, which is a reflection of the metabolic rate of the cells. The technique has not found widespread use due to the relatively high price of the instrument as well as a low signal resolution, which is particularly problematic for measurements at laboratory scale (Schügerl, 2001; Vaidyanathan et al., 1999). Off-line technologies for cell size estimation such as flow cytometry (unicellular organisms) and laser diffraction techniques (mycelial biomass) also have potential in future on-line applications. However,

the two techniques require some sample pretreatment (dilution) which may be difficult to implement on-line.

10.3.3 NIR - selection of probe type

Three probe designs are commonly used for collection of NIR spectra in situ: transmission, transflectance and reflectance (described in detail in Chapter 3). The most suitable probe design depends on the characteristics of the cultivation broth such as the concentration and size distribution of solid particles (e.g. biomass) in suspension. Table 10.4 summarizes the microbial strain, biomass range and path length used in studies employing the various probe types.

The quality of the NIR spectra is connected to the level of absorbance. At high absorptions, the light intensity is low and the signal to noise ratio is therefore also low. The probe design should be chosen such that a suitable absorbance is obtained in the wavelength interval of interest over the range of biomass concentrations. For clear broths, transmission and transflectance are most suitable, whereas for highly scattering broths, the reflectance probe is most suitable.

The transflectance probe is the most commonly used probe for cultivation monitoring (Table 10.4, Chapter 3). The path length is usually in the range 0.5-1 mm. The transmission and transflectance probes have been used for on-line measurements in cultivation broths with concentrations up to 17 g/L (Table 10.4). A short path length may constitute a problem for measurements in cultivation broths containing large biomass particles. We experienced clogging of the probe in the *S. coelicolor* fed-batch cultivations (Chapter 5). In order to avoid clogging of the probe, it is advantageous to place the probe such that the slit is not totally aligned with the flow but tilted approximately 15° (Anders Larsen, Q-Interline, personal communication).

The reflectance probe is most suited for applications with high biomass concentrations (Table 10.4). The scattering properties of the broth are not only dependent on the concentration of the biomass but also on the size of the particles. Small particles scatter the light more efficiently than large particles, and thus a higher biomass concentration is required for a good reflectance signal in cultivation broths with large biomass particles such as cell pellets.

10.4 Experimental design

For successful practical implementation of on-line sensors in the biotechnological industry, one of the most important questions to answer is how the sensors should be calibrated and where this fits in to the existing work flow from product development to large scale production. In order to obtain high quality data for calibration, there are several issues to be addressed:

- Ensuring sufficient variation in the data to calibrate a model, which covers future experiments including deviation from normal operation
- Handling of correlations between metabolite and biomass concentrations
- Acquisition of representative sensor and reference samples

Table 10.3: Overview of the prediction errors for biomass concentration obtained with the most widely used techniques for biomass monitoring.

Sensor	Prediction error (biomass range)	
	Unicellular	Filamentous
NIR	0.56-0.95 g/L (0-16 g/L) (Cimander and Mandenius, 2002; Tosi et al., 2003)	Viscosity[A] 150 cp (50-1500 cp) (Rodrigues et al., 2008)
MIR	0.845-1.73 OD (0-6 OD) (Veale et al., 2007)	-
Dielectric spectroscopy	0.6 g/L (0-7 g/L) (Dabros et al., 2009b) 1.4-1.9 g/L (7-40 g/L) (Clementschitsch et al., 2005)	2.9 ± 1.5 g/L (0-26 g/L) (Chapter 9, Petersen et al. (2010))
Multi-wavelength fluorescence	0.2-0.5 g/L (0-5 g/L) (Haack et al., 2004; Lantz et al., 2006) 0.5-2.7 g/L (0-40 g/L) (Clementschitsch et al., 2005; Ödman et al., 2009; Surribas et al., 2006a,b)	0.77-3.6 g/L (0-41 g/L) (Chapter 9, Boehl et al. (2003); Haack et al. (2007); Ödman et al. (2010))
Turbidity	R^2=0.90-0.9994 (0 -14 g/L) (Janelt et al., 2000; Kiviharju et al., 2007)	2.3 ± 1.0 g/L (0-26 g/L) (Chapter 9)

Table 10.4: Overview of microbial strains and biomass concentration range in on-line studies with transmission, transflectance and reflectance NIR probes.

Probe type	Microorganism	Biomass range	Path length	Reference
Transmission	*E. coli*	0.02-17 g/L	0.5 mm	(Arnold et al., 2002; Cimander and Mandenius, 2002)
	Vibrio cholera	0-45 ODU (at 600 nm)	0.5 mm	(Navrátil et al., 2005)
Transflectance	*S. cerevisiae*	0-14 g/L	?	(Blanco et al., 2006)
	Streptomyces coelicolor	0-8 g/L	0.5 mm	(Chapter 8, Petersen et al. (2010))
	Streptomyces clavuligerus	Viscosity: 50-1500 cp	0.5 mm	(Rodrigues et al., 2008)
	Staphylococcus, Lactobacillus, Streptococcus	0-16 g/L	1 mm	(Tamburini et al., 2003; Tosi et al., 2003)
Reflectance	*S. cerevisiae*	0-60 g/L	-	(Cavinato et al., 1990; Ge et al., 1994)
	Lactobacillus	0-18 g/L	-	(González-Vara et al., 2000; Vaccari et al., 1994)
	Unspecified industrial fungal strain	Industrial process	-	(Triadaphillou et al., 2007)

Development of new products or production strains usually takes place in research laboratories in small scale. From here the strain is transferred to pilot plant for further process development and scale up. When a desired performance has been obtained, the process is further scaled up and implemented in production. Calibration of on-line sensors for large scale operation can take place during process development and/or during production. Production development represents an obvious place for sensor calibration. Process development results in large variation in the data covering a range of operating conditions and concentrations. Ideally, the sensor could be calibrated during process development and transferred to production along with the optimized process. The main challenges of calibration during process development include the large batch-to-batch variation, which may be difficult to handle for some sensors, as well as a relatively low number of batches and short time available for calibration. In production on the other hand, one of the main challenges will be to obtain sufficient variation in the data to calibrate a model which covers a wide range of conditions and which is not based on correlations. Secondly, in large scale operation it cannot be assumed that the reactors are ideally mixed, which makes it difficult to obtain representative samples both in situ and off-line

10.4.1 Ensuring sufficient variation in the data

In order to calibrate a model which covers future experiments including deviation from normal operation, the data must contain sufficient variation. One way to obtain this is to make – small or large – changes in the protocol or operating conditions. Whereas this is a natural part of process development, it is more difficult in production, where the product has to meet the specified criteria, and where each batch may be so valuable that it is not possible to run extra experiments for the calibration of the sensor. Alternatively, small spikes of the analytes of interest can be added during the cultivation (Dabros et al., 2007). Dabros et al. (2007) used this approach for recalibration of a MIR sensor. They injected a solution designed to instantaneously raise the concentration of the five monitored analytes by the following amounts: 2.5 g/L of glucose, 2.5 g/L of ethanol, 0.2 g/L of ammonium, 0.4 g/L of phosphate, and 0.4 g/L of glycerol. This concentration shift corresponded to 1/5 of the center point of the off-line calibration set. This approach can also be used to break correlations between analytes. It is important that the spikes are sufficiently large compared to the detection limit and predictive ability of the sensor, but so small that normal operation is not disturbed, which may be difficult to obtain. The concept of perturbing an input signal to create variation for identification of model parameters is not new but has been used extensively in other areas such as chemical and electronic engineering (Ljung, 1999).

10.4.2 Handling correlations

Inherent correlations between metabolite concentrations can make it difficult to calibrate analyte specific models. Rhiel et al. (2002a) showed the effect of correlated calibration samples on the prediction performance of multivariate PLS models based on MIR spectra. They showed that if a PLS model was calibrated

on correlated samples the resulting model could only be used to predict concentrations in new samples, if a similar correlation structure was present. This was observed regardless whether or not the spectrum contained absorption peaks of the analytes of interest. However, by including semi-synthetic samples in the calibration set is was possible to calibrate much more robust models. They used two measures, a concentration space inclusion condition and an observer condition, to predict the validity of the models. The concentration space inclusion condition is fulfilled if the initial analyte concentrations, concentrations in possible added analyte mixtures and the yield coefficients in the new data set lie within the space spanned by the corresponding concentrations and yields in the calibration set (Rhiel et al., 2002a). The observer condition targets the non-absorbing species and is only fulfilled if the information about the correlation structure (initial concentration, spiked concentrations and yields) is contained in the absorbing species(Rhiel et al., 2002a). According to the authors, these two conditions can be evaluated a priori. However, in most cases, changes in yields in new experiments are not known a priori.

Different approaches have been suggested in the literature to overcome the problems of correlation such as the use of synthetic samples, semi-synthetic samples, introduction of process variation and wavelength selection (Blanco et al., 2004; Cervera et al., 2009; Finn et al., 2006; Lewis et al., 2000; McShane and Coté, 1998; Petersen et al., 2010; Rhiel et al., 2002b; Riley et al., 1998a,b; Vaidyanathan et al., 2001). Synthetic samples can be used to construct a sample set without correlations between the analytes. This approach has been used for calibration of NIR and MIR models (Blanco et al., 2004; Lewis et al., 2000; McShane and Coté, 1998; Riley et al., 1998a; Schenk et al., 2007). Schenk et al. (2007) calibrated prediction models for five different analytes based on a library of only five MIR spectra - one for each of the analytes. The models performed well for prediction of cultivation data. This method provides a fast calibration, which would be advantageous in process development where only a limited number of cultivations are run. For NIR data, which are less resolved than MIR spectra, there is general agreement that the models calibrated on synthetic samples alone do not perform well for the prediction of cultivation samples (Blanco et al., 2004; Lewis et al., 2000; McShane and Coté, 1998; Riley et al., 1998a). Instead semi-synthetic samples have been used for calibration of models based on NIR spectra and the models performed well for prediction of cultivation samples (Chapter 8, Finn et al. (2006); Petersen et al. (2010); Rhiel et al. (2002b); Riley et al. (1998b)). However, a full experiment with semi-synthetic samples will be difficult to carry out in an industrial setting. Instead, small spikes can be added to the bioreactor during operation as described above (Dabros et al., 2007). Correlations can be broken by introducing variations in the protocol and operating conditions of the cultivation process (Chapter 9, Triadaphillou et al. (2007)). For the eight batches used for calibration of biomass sensors in Chapter 9, the correlation between the biomass concentration and the glucose and ammonium concentration were 0.63 and -0.40 respectively. Finally, selection of analyte specific wavelength regions has been proposed as a means to minimize the problems of correlations in the data (Vaidyanathan et al., 2001). However, as shown by Rhiel et al. (2002a), this may not be sufficient to calibrate robust models that are also valid in new

samples with a different correlation structure.

10.4.3 Sampling

One of the biggest challenges of calibration of on-line sensors for industrial use may turn out to be sampling of the large scale bioreactors. In order to calibrate a specific model, it is of course important that the concentrations in the sample measured by the on-line sensor are the same as in the reference sample. Furthermore, in order to use this information to control the process, it is important that this measure is representative of the main parts of the bioreactor. Although it is known that concentration gradients are present in large scale bioreactors (Enfors et al., 2001; Lapin et al., 2004; Larsson et al., 1996), it is very difficult to obtain detailed information about the gradients, since these depend on multiple factors such as the viscosity, the nutrient uptake rates, the mass transfer properties etc. Thus it is difficult to determine the optimal sampling point. Furthermore, the sampling point is most often restricted by the design of the bioreactor, which only allows a certain number of sampling ports. Holm-Nielsen et al. (2008) addressed the problem of sampling in large bioreactors with a highly heterogeneous broth. In order to obtain representative samples, they measured NIR spectra and collected samples for off-line analysis through a recirculation loop. This ensured that the collected spectra corresponded to the off-line sample.

A recirculation loop may not be an option for all processes due to sterility issues. In that case the on-line sensor should be placed as close as possible to the sampling point for off-line samples. Alternatively, the on-line probes offer the unique possibility to be placed in different locations in the bioreactor, thus providing information about the gradients in the bioreactor. Multiple capacitance probes have previously been used to provide a real-time three dimensional map of viable cell mass in a 1500 hL cylindroconical bioreactor in a brewing process (Boulton et al., 2005). However, in order to use the information, it is of course important that other effects, such as differences between the probes and effects of the distance to the mixer and flow of air bubbles, are removed from the measurements. The combination of the data from on-line sensors placed at different positions in a reactor with fluid dynamic models represents an interesting topic for future work.

Finally, if the sensor is calibrated based on reference measurements, it is of course very important to choose the best possible analytical method with respect to relevance, accuracy and precision, since the sensor performance will be highly dependent on this.

10.5 Model calibration and validation

Many of the advanced on-line sensors, such as the spectroscopic sensors, provide multivariate data from which the relevant information needs to be extracted. Often, the number of variables far exceeds the number of samples. In addition, the variables are often correlated which results in an ill-posed regression problem. Standard methods such as multiple linear regression are therefore not appropriate for calibration of the sensors. Dimension reduction techniques are suitable for

solving such regression problems. The most widely used method within bioprocess monitoring is partial least squares (PLS) regression or as it is also termed projection to latent structures (Chapter 4). Related methods include principal component regression (PCR) and ridge regression. More advanced methods such as non-linear extensions of PLS (using polynomial or spline inner relations), locally weighted regression and in particular neural networks have also found widespread use in bioprocess monitoring (Ding et al., 1999; Lee et al., 2005; Wolf et al., 2001). For analysis of multiway data, methods such as parallel factor analysis (PARAFAC) and N-PLS can be used (Bro, 1996, 1997). In our experience, standard PLS performs very well for many spectroscopic data sets in combination with proper pretreatment of the data and a suitable variable selection algorithm. Furthermore, PLS is quite robust and is more easily interpretable than for example neural networks. The latter point is extremely important for approval of measurement methods in the frame of Process Analytical Technology (PAT) projects in the pharmaceutical industry and other regulated industries.

The multivariate statistical tools are very powerful and are therefore highly dependent on proper validation. This section presents a brief overview of multivariable model development practice with particular emphasis on the validation strategies. PLS is the most widely used method and the discussion is therefore based on this method but can be extended to other methods.

10.5.1 Data pretreatment

In most cases the quality of the calibration models increase greatly by pretreatment of the data prior to model calibration. Table 10.5 summarizes the most widely used pretreatments for the different sensor types. It is difficult to define general guidelines for the best pretreatment since this is highly application dependent. Thus, it is a good idea to test different pretreatments for a new application. Some of the pretreatments are more widely used than others for particular sensors, and this can be used as a guideline in the selection of a suitable pretreatment (Table 10.5). An important part of the data pretreatment is the removal of outliers. This can be done for example by evaluation of the leverage (measure of the influence of individual measurement points on the model) and the residuals (Chapter 4).

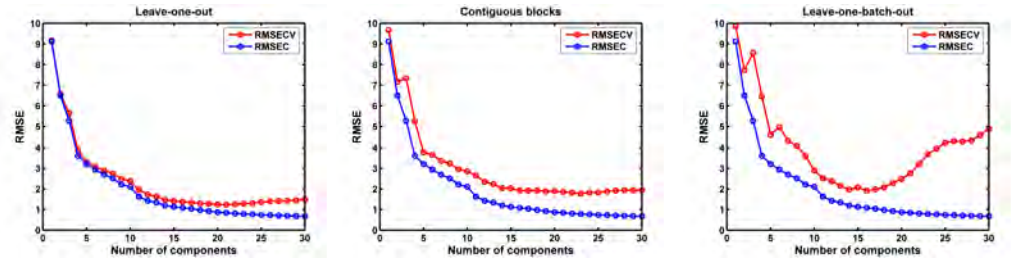
10.5.2 Variable selection

Spectroscopic methods such as NIR generate very large data sets of which only parts may contain relevant information. Most often, the quality of the model improves when the noisy regions are removed prior to model calibration. A number of different methods exist for variable selection (Table 10.5). Popular methods include manual knowledge based selection (Vaidyanathan et al., 2001), genetic algorithms (Bangalore et al., 1996; Ödman et al., 2010) and iPLS (Nørgaard et al., 2000). A thorough review of variable selection methods can be found in Xiaobo et al. (2010). We have obtained good results with a simple algorithm, in which models were calibrated for a number of wavelength regions and subsequently evaluated based on leave-one-batch-out cross-validation (Chapter 8).

In the selection of variables, one of the key points is to use a relevant criterion

Table 10.5: Overview of the pretreatments, variable selection methods and model types often used in modeling of data from NIR, MIR, Scanning DE and MWF spectroscopy.[A] Described in (Triadaphillou et al., 2007).[B] Subtraction of spectra collected at the beginning of the process.

	Pretreatment	Variable selection	Model type
NIR	-2nd derivative -MSC -SNV -1st derivative	-Manual knowledge based selection -GA -Spectral window selection[A]	-PLS -Neural networks -Locally weighted regression
MIR	-Mean centering -Baseline correction -1st derivative	-Manual knowledge based selection	-PLS
Scanning DE	-Smoothing w.r.t. time -Subtraction of measurement at background frequency -Mean centering -Batchwise standardization[B]	-Manual removal of noisy variables	-Theoretical modeling -PLS -Neural networks -Single linear regression
MWF	-Batchwise standardization[B] -Mean centering -Autoscaling	-No selection -Genetic algorithms -iPLS -Principles Variables method PLS	-PLS -PARAFAC -N-PLS -PLS on PARAFAC scores



(a) Leave-one-batch-out **(b)** Continuous Blocks **(c)** Leave-one-batch-out

Figure 10.1: Crossvalidation results for different cross-validation strategies based on the data presented in Chapter 8. (a) leave-one-sample-out; (b) contiguous blocks; (c) leave-one-batch-out.

for the evaluation of the variable subsets. For example, if there is large batch-to-batch variation, it is important to choose a criterion which reflects this (further elaborated below).

10.5.3 Calibration and internal validation

The calibration of the models can be performed easily in commercial software such as the Unscrambler (CAMO, Oslo, Norway), SIMCA (Umetrics, Umeå, Sweden) or using the PLS toolbox (Eigenvector Research Inc, WA, USA) for MATLAB. The methods are described in more detail in Chapter 4. During calibration of the PLS model, the number of components should be chosen to obtain the right balance between capturing a sufficient part of the information in the model and avoiding overfitting. This is most effectively done by choosing an appropriate (cross-)validation scheme. In our research, we have observed a high degree of batch-to-batch variation and in order to train the models to handle this we have used "leave-one-batch-out" cross-validation (Chapter 8 and 9). With a low number of batches, this is a very tough test and more moderate options can be used, for example contiguous blocks. The number of blocks can for example be chosen such that half of the batches are left out one at a time. If leave-one-sample-out cross-validation is used, there may be a danger of overfitting the data and generating a model which is not suitable for prediction of a new batch. Figure 10.1 illustrates the cross-validation errors obtained using leave-one-sample-out, contiguous blocks and leave-one-batch-out cross validation in the NIR spectra described in Chapter 8. Leave-one-sample-out cross validation provides an estimate of the prediction error which is very close to the calibration error and which may not be a realistic estimate of the prediction error in future batches. The method of contiguous blocks provides a more conservative estimate of the prediction error, while leave-one-batch-out clearly reflects the overfitting effect of using too many components in the model with an increase in the RMSE of cross-validation (Figure 10.1). The most suitable cross-validation strategy is dependent on the application and should be chosen with care.

10.5.4 External validation

For an evaluation of the predictive ability of the models, it should be tested on an independent test set. When large batch-to-batch variation is present, it is our experience that the different batches can be regarded as samples. Thus single samples from one batch are not independent and in order to validate the model on an independent test set, a new batch should be used. However, this is of course highly dependent on the particular sensor technology, the characteristics of the samples and the specific setup. Particularly for small data sets, the prediction errors depend on the particular data set used for validation. Thus too much emphasis should not be put on small improvements in prediction errors because this may reflect adjustment to that particular validation set and not a general improvement. In this thesis, we have used repeated calibration and validation in order to reflect the dependence of the prediction error and the particular validation set, and to provide a more realistic estimate of the prediction error in future batches.

Despite careful calibration and validation in order to ensure a robust model, errors in the predictions such as off-set sets may still occur in future experiments. Therefore it may be a good idea to implement a real-time sensor supervision system.

10.6 Implementation of the sensor

10.6.1 Model transfer and maintenance

In some cases the calibration model has to be transferred, e.g. to match new sensor equipment or new process facilities. This is for example the case if the monitoring is expanded from one bioreactor to cover several reactors using multiple probes, or in the case when the sensor application is moved from one production facility to another. The problems of model transfer and thus the available techniques are often specific for each sensor type. For NIR spectroscopy, methods for model transfer may be roughly divided into three approaches: (1) by ensuring a robust calibration, (2) by adjusting the calibration or (3) by adjusting the spectra (Fearn, 2001). Robustness of calibrations can be improved by careful pretreatment of the spectra or by inclusion of data from different equipment or different reactors (Roychoudhury et al., 2007). Calibration models can be adjusted by skew and bias corrections based on reference samples (Fearn, 2001). Finally, the spectra can be adjusted by different methods based on comparison of spectra collected using the different spectroscopic equipment or at different locations (Fearn, 2001).

Continued operation may be subject to drifts in the sensor or in the process, e.g. due to aging of the equipment, changes in raw materials, or optimization of the process. The calibration model should therefore be maintained in order to provide measurements of sufficiently high quality under the new conditions. The first step of model maintenance is continued monitoring of the performance, which can be achieved using some of the approaches discussed in Section 10.6.2. In practice, a model can be maintained by incorporation of new data each time they are available, and subsequent recalibration of the model based on the expanded data

set. Doing so, care should be taken to ensure that the analyte specificity is not lost in recalibration. This can for example be done by continued incorporation of samples in which the analyte concentrations are not correlated, for example samples collected during non-normal operation or spiked samples.

10.6.2 Supervision of the sensor in real-time

If the measurements of the advanced on-line sensors are subject to significant uncertainty, it is important to establish procedures for checking the sensors in real-time. A wide range of advanced methods for fault detection and diagnosis exist, which can be used to monitor both the process and the on-line sensor. These methods can be divided into quantitative model based methods, qualitative model based methods and process history based methods (Lardon et al., 2004; Venkatasubramanian et al., 2003a,b,c). A detailed review of the methods is beyond the scope of this discussion. Instead this section focuses on a few methods for supervision of on-line sensors which have been proposed in the literature for bioprocess systems.

Some of the on-line sensors, such as NIR and MIR, may be subject to offsets in the predictions caused by batch-to-batch variation or instrument drift. If the initial concentrations of the different analytes are known, it is fairly easy to correct for this offset (Rhiel et al., 2002c). However, this procedure is not useful in case of variation in the medium. Alternatively, the sensor can be recalibrated by adding small amounts of the different analytes during the process (Dabros et al., 2007). Finally, a sample can be taken for off-line analysis at the beginning of the batch. The off-line analysis can for example be performed quickly using at-line spectroscopy, which is subject to less batch-to-batch variation due to the constant setup, the possibility of referencing the instrument immediately before the measurement and the possibility to pretreat the sample. Although this solution is not optimal, it does constitute a practically feasible strategy, which can reduce the uncertainties of the on-line sensors and thus improve their applicability.

Outlier detection can be performed in real-time by evaluation of the distance of a new measurement to the other measurement points. For multivariable sensors, this can for example be done using the Mahalanobis distance or a Hotelling's T^2 test (Chapter 4).

Different on-line sensors can be used in combination to check and correct each other. Signals from different sensors, which measure the same, can be compared to see if one of the signals deviates significantly from the others (Chapter 9). A warning limit can be defined based on the Mahalanobis distance or the Hotelling's T^2 test (Chapter 4). The multivariate tools are more powerful than the corresponding univariable tools, because they are able to detect deviations in the covariance of the variables, for example a high OD combined with a low capacitance, in addition to deviations in the single variables. In well characterized systems, the measurements from the different sensors can be combined using a first principles based model (Dabros et al., 2009a). This provides a full description of the system, which ensures that the measurements are physically meaningful and which would be very useful in the interpretation of process variation.

Statistical process control has been used in combination with advanced on-line sensors for monitoring of cultivation processes (Cimander and Mandenius, 2002).

Based on prior experiments, the mean and range of the different process variables are calculated and used to evaluate new experiments in real-time (Cimander and Mandenius, 2002). Four types of trajectories may be calculated: the average process trajectory, the warning limit, action limit, and process experience limit (Cimander and Mandenius, 2002). Multivariable statistical process control is the combination of statistical process control and multivariable data analysis. By application of MSPC it is possible to utilize multivariable data sets, for example from advanced on-line sensors such as NIR and classical monitoring data. MSPC and SPC facilitates the identification of process deviations and can provide important information of the process experience limit, outside which model predictions are uncertain.

10.7 Conclusion

The successful development of a monitoring application for industrial cultivation processes requires a number of steps, which are summarized below.

- Identification of key process parameters for monitoring based on process knowledge.
- Identification of a suitable monitoring strategy based on realistic estimates of the predictive ability and range of application of the various technologies. This is highly dependent on the process characteristics such as biomass concentration, analyte concentrations, agitation, stirring etc. Furthermore, there are different requirements in process development and production:
  - Process development: the large variation in the processes and the short time line require robust sensors which are easy to calibrate. If a suitable on-line technology does not exist, at-line applications could provide the required accurate and robust measurements.
  - Production: Smaller variation in the process facilitates the use of more sensitive technologies which are able to detect small changes during the process.
- Experimental design for calibration of the sensor which ensures sufficient variation in the data and elimination of correlations.
  - Process development: The resulting data usually have sufficient variation and low correlation between concentrations. Synthetic samples or semi-synthetic samples may be used to speed up calibration.
  - Production: In production, the process usually stays within a limited range of variation. This may require the use of spiking, possibly in situ during the process, in order to break correlations and ensure sufficient variation.
- Calibration of the model. If the data are of high quality, standard PLS combined with suitable pretreatment of the data and variable selection is often adequate for calibration of the sensor.

- Validation of the sensor which mirrors the intended use of the sensor. It is often a good idea to validate the sensor on new batches in order to get a realistic estimate of the future predictive ability.

- Implementation of the sensor and a sensor supervision system in order to increase certainty in the estimates and to ensure high quality measurements in the case of process or sensor drifts. These could for example include warning limits, e.g. based on Hotelling's T^2 or SPC. In case the sensor signal exceeds the warning limit, the sensor could be checked by comparison to other sensor signals, by checking mass balances or by collection of a sample for at-line analysis.

Bibliography

Albert S, Kinley RD. 2001. Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision. Trends in Biotechnology 19:53–62.

Ansorge S, Esteban G, Schmid G. 2007. On-line monitoring of infected Sf-9 insect cell cultures by scanning permittivity measurements and comparison with off-line biovolume measurements. Cytotechnology 55:115–24.

Arnold SA, Gaensakoo R, Harvey LM, McNeil B. 2002. Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch *Escherichia coli* process. Biotechnology and Bioengineering 80:405–413.

Bachinger T, Riese U, Eriksson R, Mandenius C. 2000. Monitoring cellular state transitions in a production-scale CHO-cell process using an electronic nose. Journal of Biotechnology 76:61–71.

Bangalore AS, Shaffer RE, Small GW, Arnold MA. 1996. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least-squares regression: Application to near-infrared spectroscopy. Analytical Chemistry 68:4200–4212.

Blanco M, Peinado AC, Mas J. 2004. Analytical monitoring of alcoholic fermentation using NIR spectroscopy. Biotechnology and Bioengineering 88:536.

Blanco M, Peinado AC, Mas J. 2006. Monitoring alcoholic fermentation by joint use of soft and hard modelling methods. Analytica Chimica Acta 556:364–373.

Boehl D, Solle D, Hitzmann B, Scheper T. 2003. Chemometric modelling with two-dimensional fluorescence data for *Claviceps purpurea* bioprocess characterization. Journal of Biotechnology 105:179–188.

Boulton C, Wilson C, Peters A, Wright J. 2005. An investigation into the distribution of viable yeast mass and temperature variation in cylindroconical vessels during fermentation. Proceedings of the Congress - European Brewery Convention 30th:32/1–32/9.

Bro R. 1996. Multiway calibration. multilinear PLS. Journal of Chemometrics 10:47–61.

Bro R. 1997. PARAFAC. Tutorial and applications. Chemometrics and Intelligent Laboratory Systems 38:149–171.

Buttler T, Gorton L, Jarskog H, Marko-Varga G, Hahn-HäGerdal B, Meinander N, Olsson L. 1994. Monitoring of ethanol during fermentation of a lignocellulose hydrolysate by on-line microdialysis sampling, column liquid chromatography, and an alcohol biosensor. Biotechnology and Bioengineering 44:322–328.

Buttler T, Nilsson C, Gortona L, Marko-Vargaa G, Laurell T. 1996. Membrane characterisation and performance of microdialysis probes intended for use as bioprocess sampling units. Journal of Chromatography A 725:41–56.

Cavinato AG, Mayes DM, Ge Z, Callis JB. 1990. Noninvasive method for monitoring ethanol in fermentation processes using fiber-optic near-infrared spectroscopy. Analytical Chemistry 62:1977–1982.

Cervera AE, Petersen N, Gernaey KV, Lantz AE, Larsen A. 2009. Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. Biotechnology Progress 25:1561–1581.

Cimander C, Mandenius CF. 2002. Online monitoring of a bioprocess based on a multi-analyser system and multivariate statistical process modelling. Journal of Chemical Technology and Biotechnology 77:1157–1168.

Clarke SJ, Littleford RE, Smith WE, Goodacre R. 2005. Rapid monitoring of antibiotics using raman and surface enhanced raman spectroscopy. The Analyst 130:1019–1026.

Clementschitsch F, Bayer K. 2006. Improvement of bioprocess monitoring: development of novel concepts. Microbial Cell Factories 5.

Clementschitsch F, Kern J, Pötschacher F, Bayer K. 2005. Sensor combination and chemometric modelling for improved process monitoring in recombinant *E. coli* fed-batch cultivations. Journal of Biotechnology 120:183–197.

Correia I, Nunes A, Duarte IF, Barros A, Delgadillo I. 2005. Sorghum fermentation followed by spectroscopic techniques. Food Chemistry 90:853–859.

Dabros M, Amrhein M, Bonvin D, Marison IW, von Stockar U. 2009a. Data reconciliation of concentration estimates from mid-infrared and dielectric spectral measurements for improved on-line monitoring of bioprocesses. Biotechnology Progress 25:578–588.

Dabros M, Amrhein M, Gujral P, von Stockar U. 2007. On-line recalibration of spectral measurements using metabolite injections and dynamic orthogonal projection. Applied Spectroscopy 61:507–513.

Dabros M, Dennewald D, Currie DJ, Lee MH, Todd RW, Marison IW, von Stockar U. 2009b. Cole-Cole, linear and multivariate modeling of capacitance data for on-line monitoring of biomass. Bioprocess and Biosystems Engineering 32:161–173.

Ding Q, Small G, Arnold M. 1999. Evaluation of nonlinear model building strategies for the determination of glucose in biological matrices by near-infrared spectroscopy. Analytica Chimica Acta 384:333–343.

Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, Krüger E, Schweder T, Hamer G, O'Beirne D, Noisommit-Rizzi N, Reuss M, Boone L, Hewitt C, McFarlane C, Nienow A, Kovacs T, Trägårdh C, Fuchs L, Revstedt J, Friberg P, Hjertager B, Blomsten G, Skogman H, Hjort S, Hoeks F, Lin HY, Neubauer P, van der Lans R, Luyben K, Vrabel P, Manelius. 2001. Physiological responses to mixing in large scale bioreactors. Journal of Biotechnology 85:175 – 185.

Fearn T. 2001. Standardisation and calibration transfer for near infrared instruments: a review. Journal of Near Infrared Spectroscopy 9:229–244.

Finn B, Harvey LM, McNeil B. 2006. Near-infrared spectroscopic monitoring of biomass, glucose, ethanol and protein content in a high cell density baker's yeast fed-batch bioprocess. Yeast 23:507–517.

Ge Z, Cavinato A, Callis J. 1994. Noninvasive spectroscopy for monitoring cell density in a fermentation process. Analytical Chemistry 66:1354–1362.

Gibson T, Prosser O, Hulbert J, Marshall R, Corcoran P, Lowery P, Ruck-Keene E, Heron S. 1997. Detection and simultaneous identification of microorganisms from headspace samples using an electronic nose. Sensors and Actuators B: Chemical 44:413–422.

González-Vara A, Vaccari G, Dosi E, Trilli A, Rossi M, Matteuzzi D. 2000. Enhanced production of L-(+)-lactic acid in chemostat by *Lactobacillus casei* DSM 20011 using ion-exchange resins and cross-flow filtration in a fully automated pilot plant controlled via NIR. Biotechnology and Bioengineering 67:147–156.

Guan Y, Evans PM, Kemp RB. 1998. Specific heat flow rate: an online monitor and potential control variable of specific metabolic rate in animal cell culture that combines microcalorimetry with dielectric spectroscopy. Biotechnology and Bioengineering 58:464–477.

Haack MB, Eliasson Lantz A, Mortensen PP, Olsson L. 2007. Chemometric analysis of in-line multi-wavelength fluorescence measurements obtained during cultivations with a lipase producing *Aspergillus oryzae* strain. Biotechnology and Bioengineering 96:904–913.

Haack MB, Eliasson Lantz A, Olsson L. 2004. On-line cell mass monitoring of *Saccharomyces cerevisiae* cultivations by multi-wavelength fluorescence. Journal of Biotechnology 114:199–208.

Hantelmann K, Kollecker M, Hüll D, Hitzmann B, Scheper T. 2006. Two-dimensional fluorescence spectroscopy: A novel approach for controlling fed-batch cultivations. Journal of Biotechnology 121:410–417.

Holm-Nielsen JB, Lomborg CJ, Oleskowicz-Popiel P, Esbensen KH. 2008. On-line near infrared monitoring of glycerol-boosted anaerobic digestion processes: Evaluation of process analytical technologies. Biotechnology and Bioengineering 99:302.

Janelt G, Gerbsch N, Buchholz R. 2000. A novel fiber optic probe for on-line monitoring of biomass concentrations. Bioprocess Engineering 22:275–279.

Kiviharju K, Salonen K, Moilanen U, Eerikaeinen T. 2008. Biomass measurement online: the performance of in situ measurements and software sensors. Journal of Industrial Microbiology and Biotechnology 35:657–665.

Kiviharju K, Salonen K, Moilanen U, Meskanen E, Leisola M, Eerikaeinen T. 2007. On-line biomass measurements in bioreactor cultivations: comparison study of two on-line probes. Journal of Industrial Microbiology and Biotechnology 34:561–566.

Kornmann H, Valentinotti S, Duboc P, Marison I, von Stockar U. 2004. Monitoring and control of *Gluconacetobacter xylinus* fed-batch cultures using in situ mid-IR spectroscopy. Journal of Biotechnology 113:231–245.

Lantz AE, Jørgensen P, Poulsen E, Lindemann C, Olsson L. 2006. Determination of cell mass and polymyxin using multi-wavelength fluorescence. Journal of Biotechnology 121:544–554.

Lapa R, Lima J, Pinto I. 2003. Development of a sequential injection analysis system for the simultaneous biosensing of glucose and ethanol in bioreactor fermentation. Food Chemistry 81:141–146.

Lapin A, Müller D, Reuss M. 2004. Dynamic behavior of microbial populations in stirred bioreactors simulated with Euler-Lagrange methods: Traveling along the lifelines of single cells. Industrial and Engineering Chemistry Research 43:4647–4656.

Lardon L, Punal A, Steyer J. 2004. On-line diagnosis and uncertainty management using evidence theory - experimental illustration to anaerobic digestion processes. Journal of Process Control 14:747–763.

Larsson G, Törnkvist M, Wernersson ES, Trägårdh C, Noorman H, Enfors SO. 1996. Substrate gradients in bioreactors: Origin and consequences. Bioprocess Engineering 14:281–289.

Lee HL, Boccazzi P, Gorret N, Ram RJ, Sinskey AJ. 2004. In situ bioprocess monitoring of *Escherichia coli* bioreactions using Raman spectroscopy. Vibrational Spectroscopy 35:131–137.

Lee KI, Yim YS, Chung SW, Wei J Jand Rhee. 2005. Application of artificial neural networks to the analysis of two-dimensional fluorescence spectra in recombinant *E coli* fermentation processes. Journal of Chemical Technology and Biotechnology 80:1036–1045.

Lewis CB, McNichols RJ, Gowda A, Coté GL. 2000. Investigation of near-infrared spectroscopy for periodic determination of glucose in cell culture media in situ. Applied Spectroscopy 54:1453.

Liden H, Bachinger T, Gorton L, Mandenius C. 2000. On-line determination of non-volatile or low-concentration metabolites in a yeast cultivation using an electronic nose. Analyst 125:1123–1128.

Ljung L. 1999. System Identification. Theory for the User. Upper Saddle River, New Jersey, USA: Prentice Hall, Inc., 2nd edition.

Madrid RE, Felice CJ. 2005. Microbial biomass estimation. Critical Reviews in Biotechnology 25:97–112.

Mandenius C, Eklöv T, Lundström I. 1997. Sensor fusion with on-line gas emission multisensor arrays and standard process measuring devices in baker's yeast manufacturing process. Biotechnology and Bioengineering 55:427–438.

Markx GH, Davey CL, Kell DB. 1991. The permittistat a novel type of turbidostat. Journal of General Microbiology 137:737–744.

McShane MJ, Coté GL. 1998. Near-infrared spectroscopy for determination of glucose, lactate, and ammonia in cell culture media. Applied Spectroscopy 52:1073.

Namdev P, Alroy Y, Singh V. 1998. Sniffing out trouble: Use of an electronic nose in bioprocesses. Biotechnology Progress 14:75–78.

Navrátil M, Norberg A, Lembrén L, Mandenius C. 2005. On-line multi-analyzer monitoring of biomass, glucose and acetate for growth rate control of a *Vibrio cholerae* fed-batch cultivation. Journal of Biotechnology 115:67–79.

Nielsen J, Villadsen J, Lidén G. 2003. Bioreaction Engineering Principles. Kluwer Academic/Plenum Publishers.

Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. 2000. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Applied Spectroscopy 54:413–419.

Ödman P, Johansen C, Olsson L, Gernaey KV, Lantz A. 2010. Sensor combination and chemometric variable selection for online monitoring of *Streptomyces coelicolor* fed-batch cultivations. Applied Microbiology and Biotechnology 86:1745–1759.

Ödman P, Johansen CL, Olsson L, Gernaey KV, Lantz AE. 2009. On-line estimation of biomass, glucose and ethanol in *Saccharomyces cerevisiae* cultivations using in-situ multi-wavelength fluorescence and software sensors. Journal of Biotechnology 144:102–112.

Olsson L, Nielsen J. 1997. Online and in situ monitoring of biomass in submerged cultivations. Trends in Biotechnology 15:517–522.

Petersen N, Ödman P, Cervera AE, Eliasson Lantz A, Stocks S, Gernaey KV. 2010. In situ near infrared spectroscopy for analyte-specific monitoring of glucose and ammonium in *Streptomyces coelicolor* fermentations. Biotechnology Progress 26:263–271.

Poulsen B, Ruijter G, Panneman H, Iversen J, Visser J. 2004. Fast response filter module with plug flow of filtrate for on-line sampling from submerged cultures of filamentous fungi. Analytica Chimica Acta 510:203–212.

Rhiel M, Amrhein M, Marison I, von Stockar U. 2002a. The influence of correlated calibration samples on the prediction performance of multivariate models based on mid-infrared spectra of animal cell cultures. Analytical Chemistry 74:5227–5236.

Rhiel M, Cohen MB, Murhammer DW, Arnold MA. 2002b. Nondestructive near-infrared spectroscopic measurement of multiple analytes in undiluted samples of serum-based cell culture media. Biotechnology and Bioengineering 77:73–82.

Rhiel M, Ducommun P, Bolzonella I, Marison I, von Stockar U. 2002c. Real-time in situ monitoring of freely suspended and immobilized cell cultures based on mid-infrared spectroscopic measurements. Biotechnology and Bioengineering 77:174–185.

Riley GL, Tucker KG, Paul GC, Thomas CR. 2000. Effect of biomass concentration and mycelial morphology on fermentation broth rheology. Biotechnology and Bioengineering 68:160–172.

Riley MR, Amold MA, Murhammer DW. 1998a. Matrix-enhanced calibration procedure for multivariate calibration models with near-infrared spectra. Applied Spectroscopy 52:1339–1347.

Riley MR, Arnold MA, Murhammer DW, Walls EL, DelaCruz N. 1998b. Adaptive calibration scheme for quantification of nutrients and byproducts in insect cell bioreactors by near-infrared spectroscopy. Biotechnology Progress 14:527–533.

Rodrigues LO, Vieira L, Cardoso JP, Menezes JC. 2008. The use of NIR as a multi-parametric in situ monitoring technique in filamentous fermentation systems. Talanta 75:1356–1361.

Roychoudhury P, Harvey L, McNeil B. 2006a. At-line monitoring of ammonium, glucose, methyl oleate and biomass in a complex antibiotic fermentation process using attenuated total reflectance-mid-infrared (ATR-MIR) spectroscopy. Analytica Chimica Acta 561:218–224.

Roychoudhury P, Harvey LM, McNeil B. 2006b. The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring. Analytica Chimica Acta 571:159–166.

Roychoudhury P, Kennedy R, McNeil B, Harvey LM. 2007. Multiplexing fibre optic near infrared (NIR) spectroscopy as an emerging technology to monitor industrial bioprocesses. Analytica Chimica Acta 590:110–117.

Schenk J, Marison IW, von Stockar U. 2007. Simplified Fourier-transform mid-infrared spectroscopy calibration based on a spectra library for the on-line monitoring of bioprocesses. Analytica Chimica Acta 591:132–140.

Schenk J, Viscasillas C, Marison IW, von Stockar U. 2008. On-line monitoring of nine different batch cultures of *E-coli* by mid-infrared spectroscopy, using a single spectra library for calibration. Journal of Biotechnology 134:93–102.

Schügerl K. 2001. Progress in monitoring, modeling and control of bioprocesses during the last 20 years. Journal of biotechnology 85:149–173.

Singh R, Gernaey KV, Gani R. 2009. Model-based computer-aided framework for design of process monitoring and analysis systems. Computers and Chemical Engineering 33:22–42.

Singh R, Gernaey KV, Gani R. 2010. A software for design, analysis and validation of pat systems. Computers and Chemical Engineering 34:1108–1136.

Sivakesava S, Irudayaraj J, Demirci A. 2001. Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy. Journal of Industrial Microbiology and Biotechnology 26:185.

Suhr H, Wehnert G, Schneider K, Bittner C, Scholz T, Geissler P, Jahne B, Scheper T. 1995. In-situ microscopy for online characterization of cell-populations in bioreactors, including cell-concentration measurements by depth from focus. Biotechnology and Bioengineering 47:106–116.

Surribas A, Amigo JM, Coello J, Montesinos JL, Valero F, Maspoch S. 2006a. Parallel factor analysis combined with PLS regression applied to the on-line monitoring of *Pichia pastoris* cultures. Analytical and Bioanalytical Chemistry 385:1281–1288.

Surribas A, Geissler D, Gierse A. 2006b. State variables monitoring by in situ multi-wavelength fluorescence spectroscopy in heterologous protein production by *Pichia pastoris*. Journal of Biotechnology 124:412–420.

Tamburini E, Vaccari G, Tosi S, Trilli A. 2003. Near-infrared spectroscopy: a tool for monitoring submerged fermentation processes using an immersion optical-fiber probe. Applied Spectroscopy 57:132–138.

Torto N, Gorton L, Marko-Varga G, Emnéus J, Akerberg C, Zacchi G, Laurell T. 1997. Monitoring of enzymatic hydrolysis of starch by microdialysis sampling coupled on-line to anion exchange chromatography and integrated pulsed electrochemical detection using post-column switching. Biotechnology and Bioengineering 56:546–554.

Tosi S, Rossi M, Tamburini E, Vaccari G, Amaretti A, Matteuzzi D. 2003. Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes. Biotechnology Progress 19:1816–1821.

Treskatis S, Orgeldinger V, Wolf H, Gilles E. 1997. Morphological characterization of filamentous microorganisms in submerged cultures by on-line digital image analysis and pattern recognition. Biotechnology and Bioengineering 53:191–201.

Triadaphillou S, Martin E, Montague G, Norden A, Jeffkins P, Stimpson S. 2007. Fermentation process tracking through enhanced spectral calibration modeling. Biotechnology and Bioengineering 97:554–567.

Vaccari G, Dosi E, Campi AL, González-Vara A, Matteuzzi D, Mantovani G. 1994. A near-infrared spectroscopy technique for the control of fermentation processes: an application to lactic acid fermentation. Biotechnology and Bioengineering 43:913–917.

Vaidyanathan S, Harvey L, McNeil B. 2001. Deconvolution of near-infrared spectral information for monitoring mycelial biomass and other key analytes in a submerged fungal bioprocess. Analytica Chimica Acta 428:41–59.

Vaidyanathan S, Macaloney G, Vaughan J, McNeil B, Harvey LM. 1999. Monitoring of submerged bioprocesses. Critical Reviews in Biotechnology 19:277–316.

Veale EL, Irudayaraj J, Demirci A. 2007. An on-line approach to monitor ethanol fermentation using FTIR spectroscopy. Biotechnology Progress 23:494–500.

Venkatasubramanian V, Rengaswamy R, Kavuri SN. 2003a. A review of process fault detection and diagnosis diagnosis part I: Quantitative model-based methods. Computers and Chemical Engineering 27:293–311.

Venkatasubramanian V, Rengaswamy R, Kavuri SN. 2003b. A review of process fault detection and diagnosis part II: Quantitative model and search strategies. Computers and Chemical Engineering 27:313–326.

Venkatasubramanian V, Rengaswamy R, Kavuri SN, Yin K. 2003c. A review of process fault detection and diagnosis part III: Process history based methods. Computers and Chemical Engineering 27:327–346.

Wolf G, Almeida C J Sand Pinheiro, Correia V, Rodrigues C, Reis M, Crespo JG. 2001. Two-dimensional fluorometry coupled with artificial neural networks: A novel method for on-line monitoring of complex biological processes. Biotechnology and Bioengineering 72:297–306.

Xiaobo Z, Jiewen Z, Povey MJ, Holmes M, Hanpin M. 2010. Variables selection methods in near-infrared spectroscopy. Analytica Chimica Acta 667:14 – 32.

Xiong ZQ, Guo MJ, Guo YX, Chu J, Zhuang YP, Zhang SL. 2008. Real-time viable-cell mass monitoring in high-cell-density fed-batch glutathione fermentation by *Saccharomyces cerevisiae* T65 in industrial complex medium. Journal of Bioscience and Bioengineering 105:409–413.

Chapter 11

Summary and future perspectives

The work described in this thesis, has studied the use of various advanced sensors for monitoring of filamentous cultivations. Chapter 6 and 7 studied the use of laser diffraction for quantification of filamentous cell clumps and pellets and the relation to the rheological properties of the cultivation broth. The particle size distribution revealed multimodal particle size distributions in both *Streptomyces coelicolor* and *Aspergillus oryzae* samples. The presence of two distinct subpopulations of cell pellets was confirmed by image analysis for *S. coelicolor*. The laser scattering technique is relatively rapid (four minutes for 3 repetitive measurements) and fully automatic, which makes it an interesting alternative to image analysis for industrial use, although it does not provide the same level of detail. It was demonstrated that the utilization of the whole size distribution spectrum provided additional information of the filamentous morphology compared to just using the average size. The use of multivariable analysis aided the extraction of information from the size distribution. A PCA of the size distribution data of the *A. oryzae* samples showed that the first principal component was clearly related to the changes in morphology as the cultivation progressed. The second principal component captured the differences between the morphologies in "continuous" and "pulse-pause" batches. Finally, it was possible to calibrate models for prediction of the rheological parameters: yield stress, consistency index and the apparent viscosity of *A. oryzae* broth based on the size distribution.

In Chapter 8 the use of in situ NIR was evaluated for prediction of two important analytes: glucose and ammonium in *S. coelicolor* cultivations. In situ measurements of semi-synthetic samples were used to break the inherent correlations between glucose, ammonium and biomass concentration in the batch process. The results showed that both off-line and in situ NIR are suitable for measurement of glucose concentration in the range 1-40 g/L. For measurement of ammonium in the concentration range 40-110 mM, a large decrease in predictive ability was observed for in situ measurements compared to off-line measurements. This may be partly explained by a reduction in usable wavelength regions in the in situ NIR spectra compared to the spectra collected off-line as well as increased noise in the in situ measurements. This study thus demonstrated the specific challenges of in situ measurements in this challenging cultivation system.

A number of different advanced on-line sensors for biomass monitoring were evaluated in Chapter 9. The study included multi-wavelength fluorescence (MWF), dielectric spectroscopy (DE), on-line turbidity (OD) and software sensors. It was shown that the more advanced sensors with multivariable output did not improve predictions of dry cell weight compared to the more simple measurements of dual frequency DE, on-line turbidity and measurements of carbon dioxide evolution

rate (CER). However, the predictions of cell dry weight were greatly improved by combination of dual frequency DE, on-line OD and CER measurements. Furthermore, by using a combination of sensors it was possible to evaluate the validity of the sensor predictions in real-time, facilitating an operator intervention such as a change between the individual sensors or the collection of an off-line sample for recalibration. Thus, it was demonstrated that the on-line biomass sensors fulfills the demanding requirements of process development and can be of great industrial use in both pilot plant or production.

Based on the experience obtained through the work with the various sensor technologies, a set of guidelines for application of advanced sensors for cultivation monitoring were provided in Chapter 10. Different processes and different stages in cultivation development have different requirements. On-line sensors will be of high value during process development where it will aid an increased understanding of the dynamics of the process which will be valuable in optimization. The process development often results in a great variation in the cultivations, which is important for calibration of a robust model. Furthermore, the variations in process parameters and protocol will often minimize correlations between substrate, metabolite and biomass concentrations. The short time span of process development and the high variation in the data, require robust sensors that are easy to calibrate. Easily calibrated sensors include DE spectroscopy and on-line turbidity sensors, which require only a few batches for calibration. Other sensors such as MIR spectroscopy allow the use of synthetic samples for calibration, which will speed up the calibration. Alternatively, at-line sensors may provide a good solution in a process development environment, since they are more robust and easy to calibrate. In production, on-line sensors will be of great value for ensuring optimal control of the process. Moreover, they can provide early detection of process deviations and as well as their cause, e.g. differences in raw materials or contamination. For calibration in a production process, spiking schemes can be employed in order to introduce variation in the data and break correlations. Furthermore, it is very important to employ a suitable sampling strategy in the large scale bioreactors.

In conclusion, the successful application of advanced sensor technology in industrial cultivations depends on the identification of key process parameters and the matching to a suitable monitoring technology as well as the design of a calibration and validation procedure which fits into the industrial work flow. This work has provided a critical evaluation of the predictive ability of some the most promising technologies such as NIR, MWF and DE spectroscopy. Furthermore, it has highlighted the importance of the experimental design for sensor calibration to ensure sufficient variation in the data and to eliminate correlations between analytes. In future work, it would be interesting to look at a better utilization of the large amounts of information from the advanced sensors without necessarily correlating it to known analytes. Furthermore, it would be interesting to use the in situ sensors for 3-dimensional characterization of large scale reactors. The full potential of the advanced sensors will most likely only be realized once they are implemented in industrial processes, which will allow collection of data far exceeding the extent of academic studies.

Appendix A

Overview of near infrared spectroscopy studies included in review

Table A.1: Overview of type of fermentations, modeled analytes, wavelength selection, type of modeling and results for the papers cited in the review of NIR spectroscopy in Chapter 3.

Authors and year	**Production organism**	**Sampling**	**NIR mode**	**Analytes**	**Wavelength Intervals(nm)**	**Type of modeling**	**Results**
Arnold et al. (2000)	*Streptomyces fradiae*	At-line	Transmission	Methyl oleate Glucose Glutamate Ammonium	1600-1800 2083-2342 2222-2353 2100-2300	2D PLS 4LV 2D PLS 6LV 2D PLS 3LV 2D PLS 3LV	SEP=0.4203 g/L SEP=0.5925 g/L SEP=0.2532 g/L SEP=0.0161 g/L
Arnold et al. (2001)	*Streptomyces fradiae*	At-line	Reflectance	Tylosin	1600-1800	2D PLS 5-8LV	SEP=0.3938 a.u.
Arnold et al. (2002)	*Escherichia coli*	At-line On-line (in situ)	Transmission	Biomass	At-line 2050-2350 On-line 1600-1800	At-line 2D PLS On-line 2D PLS	At-line SEP=1.45 g/L On-line SEP=1.39 g/L
Arnold et al. (2003)	CHO-cells	On-line (in situ)	Transmission	Glucose Lactate Glutamine Ammonia	1650-1750, 2260-2290 400-2200 1430-1470 2300-2350	2D PLS 9LV SNV 2D PLS 8LV SNV 2D PLS 10LV SNV 2D PLS	SEP=0.072 g/L SEP=0.0144 g/L SEP=0.308 mM SEP=0.036 mM
Cavinato et al. (1990)	*S. cerevisiae*	On-line (ex situ)	Reflectance	Ethanol	905	2D MLR	SEP=0.4%(w/w)
Cimander et al. (2002)	*Lactobacillus bulgaricus* *Streptococcus thermophilus*	On-line (in situ) Sensor fusion EN and SBM	Reflectance	pH Lactose Galactose Lactate State	1402-1408	2D ANN	SEP=0.243 SEP=0.301%(w/w) SEP=0.066%(w/w) SEP=0.062%(w/w) 96.1%correct class
Cimander and Mandenius (2002)	*Escherichia coli*	On-line (in situ)	Transmission	Biomass Tryptophan	1390,1564 1674,1826	-	RMSEP=0.56 g/L rRMSEP=8%min

Continued on next page

Authors and year	Production organism	Sampling	NIR mode	Analytes	Wavelength Intervals(nm)	Type of modeling	Results
Crowley et al. (2005)	*Pichia pastoris*	At-line	Transmission (low biom. conc.)	Transmission Glycerol	Transmission 1696-1706, 2072-2098, 2264-2284	Transmission MSC 2D PLS 5LV	Transmission SEP=2.01 g/L
				Methanol	1160-1230, 2200-2450	SNV 2D PLS 3LV	SEP=0.24 g/L
			Reflectance (high biom. conc.)	Biomass	944	2D SLR	SEP=1.43 g/L
				Reflectance Methanol	Reflectance 1180-1220, 2250-2450	Reflectance SNV 2D PLS 3LV	Reflectance SEP=0.22 g/L
				Protein	400-2500	2D PLS 4LV	SEP=0.06 g/L
Ferreira et al. (2005)	*Streptomyces clavuligerus*	At-line	Reflectance	Clavulanic acid	GA or PLS-bootstrap	1D PLS 7-8LV	rRMSEP=12%min
Ferreira and Menezes (2006)	*Streptomyces clavuligerus*	On-line (in situ)	Reflectance	Identification of fermentation state transitions	833-2380	SS-2DCoS	-
Finn et al. (2006)	*S. cerevisiae*	At-line	Transmission	Biomass	910-930	2D PLS	SEP=2.45 g/L
				Ethanol	1678-1700, 2254-2272, 2294-2312	2D PLS	SEP=0.22 g/L
				Glucose	2258-2288	2D PLS	SEP=0.80 g/L
				Intracel. protein	750-780, 900-950	2D PLS	SEP=0.98 g/L
Ge et al. (1994)	*S. cerevisiae*	On-line (ex situ)	Reflectance	Biomass	700-1100	LWR	SEP=1.6 g/L
González-Vara et al. (2000)	*Lactobacillus*	On-line (ex situ)	Reflectance	Glucose Lactic acid Biomass	Described in Vaccari et al. (1994)	Described in Vaccari et al. (1994)	Described in Vaccari et al. (1994)
Hagman and Sivertsson (1998)	CHO-cells	On-line (both in- and ex situ)	Transmission	Glucose Lactate Ammonia Biomass Viability	-	-	-

Continued on next page

Authors and year	Production organism	Sampling	NIR mode	Analytes	Wavelength Intervals(nm)	Type of modeling	Results
Holm-Nielsen et al. (2008)	Thermophilic anaerobic digestion of manure	On-line (ex situ)	Trans-flectance	Glycerol Acetic acid Propanoic acid Iso-but. acid Butanoic acid Iso-valeric acid Valeric acid Volatile Fat.Ac.	960-1600	PLS 2LV PLS 3LV PLS 1LV PLS 4LV PLS 5LV PLS 6LV PLS 3LV PLS 4LV	RMSEP=3.037 g/L RMSEP=1.476 g/L RMSEP=1.364 g/L RMSEP=0.026 g/L RMSEP=0.272 g/L RMSEP=0.056 g/L RMSEP=0.024 g/L RMSEP=2.095 g/L
Lewis et al. (2000)	Fibroblast cells	On-line (in situ)	Transmission	Glucose	2040-2380	Raw PLS 10 LV	SEP=0.148 g/L
McShane and Coté (1998)	Fibroblast cells	At-line	Transmission	Glucose Lactate Ammonium	2060-2360	PLS 12LV PLS 10LV PLS 12LV	SEP=0.37 mM SEP=0.40 mM SEP=0.24 mM
Navrátil et al. (2004)	Yogurt *Lactococcus thermophilus Lactobacillus bulgaricus* Filmjölk *Lactococcus lactis, Lactococcus cremoris, Lactococcus diacetylactis, Leuconostoc cremoris*	On-line (in situ)	Reflectance	pH Titratable acidity	700-1900	Yogurt PLS 6LV Filmjölk PLS 5LV Yogurt PLS 6LV Filmjölk PLS 5LV	Yogurt SEP=0.17 Filmjölk SEP=0.22 Yogurt SEP=6.6^{o}T Filmjölk SEP=7.7^{o}T

Continued on next page

Authors and year	Production organism	Sampling	NIR mode	Analytes	Wavelength Intervals(nm)	Type of modeling	Results
Navrátil et al. (2005)	*Vibrio cholerae*	On-line (in situ)	Transmission	Biomass Glucose Acetate Cholera toxin	904-1406 904-1154 1072-1742 1000-2500	2D PLS 4LV 2D PLS 3LV 2D PLS 5LV 2D PLS 6LV	SEP=0.2 g/L SEP=0.26 g/L SEP=0.28 g/L SEC=0.02 g/L
Petersen et al. (2010)	*Streptomyces coelicolor*	On-line (in situ) Off-line	On-line Transflectance Off-line Transmission	Glucose Ammonium	Selection of intervals in 1100-2200	On-line 2D PLS Off-line 2D PLS On-line (SNV) 1D PLS Off-line (SNV) 1D PLS	On-line RMSEP=2.0±1.1g/L Off-line RMSEP=1.5±0.5g/L On-line RMSEP=11±7.6 mM Off-line RMSEP=4±1.2 mM
Rhiel et al. (2002)	PC-3 human prostate cancer cells	At-line	Transmission	Glucose Lactate Glutamine Ammonia	2247-2326 2299-2315 2179-2198 2119-2208	PLS 5LV PLS 3LV PLS 3LV PLS 6LV	SEP=0.82 mM SEP=1.01 mM SEP=0.89 mM SEP=0.76 mM
Riley et al. (1997)	Insect cells	Off-line	Transmission	Glucose Glutamine	2247-2326 2174-2247	PLS 4LV PLS 8LV	SEP=1.46 mM SEP=0.51 mM
Riley et al. (1998a)	Insect cells	Off-line	Transmission	Glucose Glutamine	2183-2381 2146-2381	PLS 7LV PLS 10LV	SEP=1.1 mM SEP=0.7 mM
Riley et al. (1998b)	Insect cells (*Spodoptera frugiperda* Sf-9)	Off-line	Transmission	Alanine Glucose Glutamine Leucine	2217-2353 2137-2336 2128-2326 2188-2326	PLS 9LV PLS 10LV PLS 11LV PLS 11LV	SEP=1.4 mM SEP=1.0 mM SEP=1.1 mM SEP=0.31 mM
Riley et al. (1999)	- (mammalian cell culture media)	Off-line + Simulated spectra	Transmission	Ammonia Glucose Glutamate Glutamine Lactate	Selection of intervals in 2000-2500	-	SEP=0.7-0.8 mM SEP=0.7-0.9 mM SEP=0.7-0.8 mM SEP=0.38-0.44 mM SEP=0.45-0.55 mM

Continued on next page

Authors and year	Production organism	Sampling	NIR mode	Analytes	Wavelength Intervals(nm)	Type of modeling	Results
Riley et al. (2001)	- (synthetic cell culture medium)	Off-line	Transmission	Glucose Lactate Ammonia Pyruvate Glutamine 14 Amino acids	Selection of intervals in 2000-2500	-	Percent errors ranging 3-37%
Rodrigues et al. (2008)	*Streptomyces clavuligerus*	On-line (in situ)	Trans-flectance	C-source N-source Viscosity Clavulanic acid	Selection by GA in 909-2500	1D PLS 1LV 1D PLS 1LV - PLS 2LV SNV PLS 6LV	RMSEP=0.4%(w/w) RMSEP=0.01%(w/w) RMSEP=150 cp RMSEP=0.03%(w/w)
Roychoudhury et al. (2007)	CHO cells	On-line (in situ)	Trans-flectance	Glucose Lactate	1333-1640 1835-2224 1121-1399	Single probe MSC PLS 4LV Multiple probes MSC PLS 6LV Single probe MSC PLS 4LV Multiple probes MSC PLS 7LV	Single probe SEP=0.2 g/L Multiple probes SEP=0.2 g/L Single probe SEP=0.1 g/L Multiple probes SEP=0.1 g/L
Tamburini et al. (2003)	*Staphylococcus* *Lactobacillus*	On-line (in situ)	Transflection	*Staphylococcus* Glucose Lactic acid Acetic acid Biomass *Lactobacillus* Glucose Lactic acid Acetic acid Biomass	700-1800	2D PLS	*Staphylococcus* SEP=2.62 g/L SEP=1.37 g/L SEP=0.63 g/L SEP=0.85 g/L *Lactobacillus* SEP=2.27 g/L SEP=0.86 g/L SEP=0.48 g/L SEP=0.55 g/L

Continued on next page

Authors and year	**Production organism**	**Sampling**	**NIR mode**	**Analytes**	**Wavelength Intervals(nm)**	**Type of modeling**	**Results**
Tosi et al. (2003)	*Staphylococcus xylosus* *Lactobacillus fermentum* *Streptococcus thermophylus*	On-line (in situ)	Transflection	*S. xylosus* Glucose Lactic acid Acetic acid Biomass *L. fermentum* Glucose Lactic acid Acetic acid Biomass *S. thermophyl.* Glucose Lactic acid Biomass	700-1800	2D PLS	*S. xylosus* SEP=2.6 g/L SEP=1.4 g/L SEP=0.6 g/L SEP=0.8 g/L *L. fermentum* SEP=1.8 g/L SEP=1.6 g/L SEP=0.5 g/L SEP=1.0 g/L *S. thermophyl.* SEP=3.6 g/L SEP=1.5 g/L SEP=0.7 g/L
Triadaphillou et al. (2007)	A fungus (not specified)	On-line (ex situ)	Reflectance	Product Phosphate Ammonium	Selection of intervals by SWS in 950-1700	1D PLS	-
Vaccari et al. (1994)	*Lactobacillus casei*	On-line (ex situ)	Reflectance	Lactic acid Glucose Biomass	-	-	σ=1.64 g/L σ=2.09 g/L σ=0.73 g/L
Vaidyanathan et al. (1999)	*E. coli* *Streptomyces fradiae* *Penicillium chrysogenum* *Aspergillus niger* *Aureobasidium pullulans*	Off-line	Transmission Reflectance	Biomass	1650-1800 2270, 2310, 2350	2D	-

Continued on next page

Authors and year	Production organism	Sampling	NIR mode	Analytes	Wavelength Intervals(nm)	Type of modeling	Results
Vaidyanathan et al. (2000)	*Streptomyces fradiae*	At-line	Transmission Reflectance	Oil	- (confidential)	Transmission 2D SLR Reflectance 2D MLR	Transmission SEP_{nor}=4.27% Reflectance SEP_{nor}=3.67%
				Tylosin		Transmission 2D PLS 4LV Reflectance 2D PLS 5LV	Transmission SEP_{nor}=3.69% Reflectance SEP_{nor}=8.42%
Vaidyanathan et al. (2001a)	*Streptomyces fradiae*	At-line	Transmission	Methyl oleate	930, 1600-1800, 2080-2360	2D PCA	-
				Glucose	900, 1690, 1730, 2100, 2270		
				Glutamate	900, 1690, 1730		
				Ammonium	900, 1690		
				Biomass	930, 1690		
				Tylosin	930, 1690, 2320		
Vaidyanathan et al. (2001b)	*Penicillium chrysogenum*	At-line	Transmission	Whole broth	Whole broth	Whole broth	Whole broth
				Total sugars	2100-2350	2D PLS 3LV	SEP=2.5 g/L
				Ammonium	2100-2200	2D PLS 3LV	SEP=0.17 g/L
				Biomass	1600-1800	2D PLS 4LV	SEP=0.74 g/L
				Filtrate samples	Filtrate samples	Filtrate samples	Filtrate samples
				Penicillin	1600-1800	2D PLS 4LV	SEP=15.1 g/L
				Extracel. prot.	1120-1300, 1450-1820, 2040-2340	2D PLS 5LV	SEP=33.0 g/L
Vaidyanathan et al. (2001c)	*Penicillium chrysogenum*	At-line	Transmission	Total sugars (lactose and sucrose)	2100-2350	2D PLS 3LV	SEP=2.5 g/L
				Ammonium	2100-2200	2D PLS 3LV	SEP=0.17 g/L
				Biomass	1600-1800	2D PLS 4LV	SEP=0.74 g/L

Continued on next page

Authors and year	Production organism	Sampling	NIR mode	Analytes	Wavelength Intervals(nm)	Type of modeling	Results
Vaidyanathan et al. (2003)	*Streptomyces fradiae*	Off-line	Transmission Reflectance	Analysis of different biomass morphologies	400-800 800-1000 1600-1750 2100-2350	PCA	-

[1] Abbreviations: Wavelength selection: SWS (spectral window selection), GA (genetic algorithms). Pretreatments: 1D (first derivative), 2D (second derivative), SNV (standard normal variate), MSC (multiplicative scatter correction). Note that scaling and mean centering are not included. Modeling: PLS (partial least squares), LV (latent variables), PCA (principal component analysis), MLR (multiple linear regression), SLR (simple linear regression), ANN (artificial neural networks), SS-2DCoS (sample-sample two-dimensional correlation spectroscopy), LWR (locally weighted regression). Results: SEP (standard error of prediction), SEPnor (normalized SEP), RMSEP (root mean squared error of prediction), rRMSEP (relative RMSEP), SEC (standard error of calibration), (errors) (standard deviation of the errors), a.u. (arbitrary units), %min (minimum error, i.e. RMSEP divided by the maximum concentration in the range). [2] When several models are tested, only the one resulting in the lowest SEP or RMSEP is shown.

Bibliography

Arnold SA, Crowley J, Vaidyanathan S, Matheson L, Mohan P, Hall J, Harvey LM, McNeil B. 2000. At-line monitoring of a submerged filamentous bacterial cultivation using near-infrared spectroscopy. Enzyme and Microbial Technology 27:691–697.

Arnold SA, Crowley J, Woods N, Harvey LM, McNeil B. 2003. In-situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation. Biotechnology and Bioengineering 84:13.

Arnold SA, Gaensakoo R, Harvey LM, McNeil B. 2002. Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch *Escherichia coli* process. Biotechnology and Bioengineering 80:405–413.

Arnold SA, Matheson L, Harvey LM, McNeil B. 2001. Temporally segmented modelling: a route to improved bioprocess monitoring using near infrared spectroscopy? Biotechnology Letters 23:143–147.

Cavinato AG, Mayes DM, Ge Z, Callis JB. 1990. Noninvasive method for monitoring ethanol in fermentation processes using fiber-optic near-infrared spectroscopy. Analytical Chemistry 62:1977–1982.

Cimander C, Carlsson M, Mandenius CF. 2002. Sensor fusion for on-line monitoring of yoghurt fermentation. Journal of Biotechnology 99:237–248.

Cimander C, Mandenius CF. 2002. Online monitoring of a bioprocess based on a multi-analyser system and multivariate statistical process modelling. Journal of Chemical Technology and Biotechnology 77:1157–1168.

Crowley J, Arnold S, Wood N, Harvey L, McNeil B. 2005. Monitoring a high cell density recombinant *Pichia pastoris* fed-batch bioprocess using transmission and reflectance near infrared spectroscopy. Enzyme and Microbial Technology 36:621–628.

Ferreira AP, Menezes JC. 2006. Monitoring a complex medium fermentation with sample-sample two-dimensional FT-NIR correlation spectroscopy. Biotechnology Progress 22:866–872.

Ferreira AP, Vieira LM, Cardoso JP, Menezes JC. 2005. Evaluation of a new annular capacitance probe for biomass monitoring in industrial pilot-scale fermentations. Journal of Biotechnology 116:403–409.

Finn B, Harvey LM, McNeil B. 2006. Near-infrared spectroscopic monitoring of biomass, glucose, ethanol and protein content in a high cell density baker's yeast fed-batch bioprocess. Yeast 23:507–517.

Ge Z, Cavinato A, Callis J. 1994. Noninvasive spectroscopy for monitoring cell density in a fermentation process. Analytical Chemistry 66:1354–1362.

González-Vara A, Vaccari G, Dosi E, Trilli A, Rossi M, Matteuzzi D. 2000. Enhanced production of L-(+)-lactic acid in chemostat by *Lactobacillus casei* DSM 20011 using ion-exchange resins and cross-flow filtration in a fully automated pilot plant controlled via NIR. Biotechnology and Bioengineering 67:147–156.

Hagman A, Sivertsson P. 1998. The use of NIR spectroscopy in monitoring and controlling bioprocesses. Process Control and Quality 11:125–128.

Holm-Nielsen JB, Lomborg CJ, Oleskowicz-Popiel P, Esbensen KH. 2008. Online near infrared monitoring of glycerol-boosted anaerobic digestion processes: Evaluation of process analytical technologies. Biotechnology and Bioengineering 99:302.

Lewis CB, McNichols RJ, Gowda A, Coté GL. 2000. Investigation of near-infrared spectroscopy for periodic determination of glucose in cell culture media in situ. Applied Spectroscopy 54:1453.

McShane MJ, Coté GL. 1998. Near-infrared spectroscopy for determination of glucose, lactate, and ammonia in cell culture media. Applied Spectroscopy 52:1073.

Navrátil M, Cimander C, Mandenius CF. 2004. On-line multisensor monitoring of yogurt and filmjölk fermentations on production scale. Journal of Agricultural and Food Chemistry 52:415–420.

Navrátil M, Norberg A, Lembrén L, Mandenius C. 2005. On-line multi-analyzer monitoring of biomass, glucose and acetate for growth rate control of a *Vibrio cholerae* fed-batch cultivation. Journal of Biotechnology 115:67–79.

Petersen N, Ödman P, Cervera AE, Eliasson Lantz A, Stocks S, Gernaey KV. 2010. In situ near infrared spectroscopy for analyte-specific monitoring of glucose and ammonium in *Streptomyces coelicolor* fermentations. Biotechnology Progress 26:263–271.

Rhiel M, Cohen MB, Murhammer DW, Arnold MA. 2002. Nondestructive near-infrared spectroscopic measurement of multiple analytes in undiluted samples of serum-based cell culture media. Biotechnology and Bioengineering 77:73–82.

Riley MR, Amold MA, Murhammer DW. 1998a. Matrix-enhanced calibration procedure for multivariate calibration models with near-infrared spectra. Applied Spectroscopy 52:1339–1347.

Riley MR, Arnold MA, Murhammer DW, Walls EL, DelaCruz N. 1998b. Adaptive calibration scheme for quantification of nutrients and byproducts in insect cell bioreactors by near-infrared spectroscopy. Biotechnology Progress 14:527–533.

Riley MR, Crider HM, Nite ME, Garcia RA, Woo J, Wegge RM. 2001. Simultaneous measurement of 19 components in serum-containing animal cell culture media by fourier transform near-infrared spectroscopy. Biotechnology Progress 17:376–378.

Riley MR, Okeson CD, Frazier BL. 1999. Rapid calibration of near-infrared spectroscopic measurements of mammalian cell cultivations. Biotechnology Progress 15:1133–1141.

Riley MR, Rhiel M, Zhou X, Arnold MA, Murhammer DW. 1997. Simultaneous measurement of glucose and glutamine in insect cell culture media by near infrared spectroscopy. Biotechnology and Bioengineering 55:11.

Rodrigues LO, Vieira L, Cardoso JP, Menezes JC. 2008. The use of NIR as a multi-parametric in situ monitoring technique in filamentous fermentation systems. Talanta 75:1356–1361.

Roychoudhury P, Kennedy R, McNeil B, Harvey LM. 2007. Multiplexing fibre optic near infrared (NIR) spectroscopy as an emerging technology to monitor industrial bioprocesses. Analytica Chimica Acta 590:110–117.

Tamburini E, Vaccari G, Tosi S, Trilli A. 2003. Near-infrared spectroscopy: a tool for monitoring submerged fermentation processes using an immersion optical-fiber probe. Applied Spectroscopy 57:132–138.

Tosi S, Rossi M, Tamburini E, Vaccari G, Amaretti A, Matteuzzi D. 2003. Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes. Biotechnology Progress 19:1816–1821.

Triadaphillou S, Martin E, Montague G, Norden A, Jeffkins P, Stimpson S. 2007. Fermentation process tracking through enhanced spectral calibration modeling. Biotechnology and Bioengineering 97:554–567.

Vaccari G, Dosi E, Campi AL, González-Vara A, Matteuzzi D, Mantovani G. 1994. A near-infrared spectroscopy technique for the control of fermentation processes: an application to lactic acid fermentation. Biotechnology and Bioengineering 43:913–917.

Vaidyanathan S, Arnold A, Matheson L, Mohan P, Macaloney G, McNeil B, Harvey LM. 2000. Critical evaluation of models developed for monitoring an industrial submerged bioprocess for antibiotic production using near-infrared spectroscopy. Biotechnology Progress 16:1098–1105.

Vaidyanathan S, Arnold SA, Matheson L, Mohan P, McNeil B, Harvey LM. 2001a. Assessment of near-infrared spectral information for rapid monitoring of bioprocess quality. Biotechnology and Bioengineering 74:376–388.

Vaidyanathan S, Harvey L, McNeil B. 2001b. Deconvolution of near-infrared spectral information for monitoring mycelial biomass and other key analytes in a submerged fungal bioprocess. Analytica Chimica Acta 428:41–59.

Vaidyanathan S, Macaloney G, Harvey L, McNeil B. 2001c. Assessment of the structure and predictive ability of models developed for monitoring key analytes in a submerged fungal bioprocess using near-infrared spectroscopy. Applied Spectroscopy 55:444–453.

Vaidyanathan S, McNeil B, Macaloney G. 1999. Fundamental investigations on the near-infrared spectra of microbial biomass as applicable to bioprocess monitoring. The Analyst 124:157–162.

Vaidyanathan S, White S, Harvey LM, McNeil B. 2003. Influence of morphology on the near-infrared spectra of mycelial biomass and its implications in bioprocess monitoring. Biotechnology and Bioengineering 82:715–724.

Appendix B

Matrix notation for efficient development of first-principles models within PAT applications: integrated modeling of antibiotic production with *Streptomyces coelicolor*

ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Matrix Notation for Efficient Development of *First-Principles* Models Within PAT Applications: Integrated Modeling of Antibiotic Production With *Streptomyces coelicolor*

Gürkan Sin,[1] Peter Ödman,[2] Nanna Petersen,[1] Anna Eliasson Lantz,[2] Krist V. Gernaey[1]

[1]Department of Chemical and Biochemical Engineering, Center for Bioprocess Engineering, Technical University of Denmark, Building 229, DK-2800 Kgs. Lyngby, Denmark; telephone: +45-45252960; fax: +45-45932906; e-mail: gsi@kt.dtu.dk

[2]Center for Microbial Biotechnology, Biocentrum-DTU, Technical University of Denmark, Lyngby, Denmark

Received 30 October 2007; revision received 4 February 2008; accepted 11 February 2008

Published online 7 March 2008 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21869

ABSTRACT: A matrix notation coupled to macroscopic principles is introduced as a means to develop first-principles models in an efficient and structured way within PAT applications. The notation was evaluated for developing an integrated biological, chemical (pH modeling) and physical (gas–liquid exchange) model for describing antibiotic production with *Streptomyces coelicolor* in batch fermentations. The model provided statistically adequate fits to all the monitored macroscopic biological, chemical and physical data of the process, except the phosphate uptake dynamics. This phosphate discrepancy is hypothesized to result from the internal storage of phosphate as polyphosphate prior to the exponential growth phase. The antibiotic production was associated with the stationary phase and its kinetics was adequately described using a modified Luedeking–Piret equation. Further, the maintenance was best described by employing a combination of Pirt and Herbert models, a result that was supported by a model-based hypothesis testing. Overall the process knowledge currently incorporated in the model is believed to be useful both for process optimization purposes and for further testing of hypotheses aiming at improving the mechanistic understanding of antibiotic production with *S. coelicolor*. Last but not least, the matrix notation is believed to be a promising supporting tool for efficient development and communication of complex dynamic models within a PAT framework.

Biotechnol. Bioeng. 2008;101: 153–171.

KEYWORDS: antibiotics; *S. coelicolor*; modeling; PAT; parameter estimation; pH

Correspondence to: G. Sin

This article contains Supplementary Material available at http://www.interscience.wiley.com/jpages/0006-3592/suppmat.

Introduction

In the pharmaceutical industry, a significant gap was recognized between the advanced scientific practice used in research and development on the one hand, and the more conservative and traditional techniques used in the manufacturing of pharmaceuticals on the other hand. The recognition of this gap was one of the most important factors leading to the development of the FDA Process Analytical Technology (PAT) initiative which aims at stimulating the implementation of innovative and risk-based approaches in the pharmaceutical industry (FDA, 2004). The development of rapid analytical technology for replacement of traditional laboratory testing for process monitoring and control has been the main focus in many past PAT applications (Junker and Wang, 2006; Maes and Van Liedekerke, 2006). The mechanistic process modeling, however, started only recently to receive attention within PAT applications (Dassau et al., 2006). Within the PAT context, which is indeed the continuous quest for increased process understanding, the employment of an integrated systems approach (not only multivariate statistical methods but also mechanistic modeling) is expected to help better reach the goals of PAT applications. Within such an integrated approach, a mechanistic model can be used as a means of summarizing and structuring available process understanding.

Fermentation processes are essential for the production of many antibiotics, vaccines and therapeutic proteins. Since the first quantitative description of bacterial growth kinetics was postulated in the beginning of the 19th century, the mechanistic modeling activities have seen a continuous growth, which resulted in the development of a wealth of modeling techniques: unstructured (Esener et al., 1983), structured, segregated, cybernetic, morphological (Nielsen

and Villadsen, 1992), verbal, black-box, stochastic, distributed (Thilakavathi et al., 2007), metabolic flux analysis (MFA), flux balance analysis (FBA) and kinetic modeling for metabolic control analysis (Teusink and Smid, 2006). The main purpose of most model applications has been to test hypotheses and thereby increase the fundamental understanding of cell metabolism at the research level, which is subsequently transferred to production scale to improve process design, control and optimization.

When opting for the use of a mechanistic model on a fermentation process within a PAT application, selection of an appropriate model type among the many existing categories developed so far will be an important step. This study takes the view that first-principles modeling incorporating unstructured (and simplified structured) models coupled with macroscopic mass balances will be valuable for at least process design and control applications. To this end, the main aim of this study is to introduce a systematic and efficient methodology based on matrix notation to facilitate the development of first-principles models within PAT applications. Such a modeling methodology is deemed particularly useful, *if not necessary*, to facilitate communication and knowledge transfer between the members of the multi-disciplinary team—researchers, engineers, operators, managers—typically involved in the development and implementation of a PAT system.

This manuscript is organized as follows: First the model matrix notation principles are introduced and illustrated using a simple Monod–Herbert aerobic bacterial growth model. Second, to illustrate and evaluate the proposed methodology, the notation is then used to develop an integrated biological, chemical and physical (mainly gas–liquid exchange) model for *Streptomyces coelicolor* fermentations for the production of actinorhodin and undecylprodigiosin antibiotics. Subsequently, the integrated model is evaluated using batch fermentation data containing biological, chemical and off-gas measurements, and results are discussed critically.

A Matrix Notation for *First-Principles* Models

The matrix notation was chosen as the basis of the proposed methodology, since it allows a compact and visually appealing description of complex mathematical models (Henze et al., 2000; Noorman et al., 1991). For description of a system with a number of components m and a number of processes n, the matrix formalism requires definition of the stoichiometric matrix, $\mathbf{S}$, the process rates vector, $\boldsymbol{\rho}$, and the component conversion rate vector, $\mathbf{r}$. The component conversion rate vector, $\mathbf{r}$ is given by

$$\mathbf{r}_{m\times 1} = \mathbf{S}'_{n\times m} \cdot \boldsymbol{\rho}_{n\times 1} \tag{1}$$

The overall component conversion rate vector is then coupled to a general mass balance equation around a system boundary. This finally provides a set of ordinary differential equations (ODE) (a vector of size m) describing dynamic system behavior, which in its general form for a reactor with one inlet (feed stream) and one outlet (product stream) looks like:

$$\frac{\mathrm{d}(V_L \cdot \mathbf{C})}{\mathrm{d}\mathbf{t}} = Q \cdot (\mathbf{C}_{in} - \mathbf{C}) - \mathbf{r} \cdot V_L \tag{2}$$

In the above equation, $\mathbf{C}_{in}$ is a vector of concentrations in the feed stream (size m), Q is the flow rate of the inlet stream (L/h), and V_L is the volume of the reactor (liquid phase). For a continuous system with constant volume, Equation (2) can then for example be rewritten as

$$\frac{\mathrm{d}\mathbf{C}}{\mathrm{d}\mathbf{t}} = \frac{Q}{V_L} \cdot (\mathbf{C}_{in} - \mathbf{C}) - \mathbf{r} \tag{3}$$

The model matrix presentation is illustrated in Table I using a simple Monod–Herbert (1958) model for growth and endogenous respiration. In Table I, the system components considered are substrate, oxygen and biomass, hence m is 3. These are given in the first row of Table I. The two processes considered are growth and decay, hence n is 2. The processes are given in the first column of Table I. The stoichiometric matrix, $\mathbf{S}$, has the dimension $\mathbf{n} \times \mathbf{m}$ and contains the stoichiometric coefficients for the components that are involved in each process. The $\mathbf{S}$ matrix corresponds to the grey shaded area and is situated in the main body of Table I. The process rates, $\boldsymbol{\rho}$, are shown in the last column of Table I.

Table I. Matrix presentation of a Monod–Herbert aerobic biomass growth model.

Component, $i\rightarrow$	C_1	C_2	C_3	Rates, ρ_j
Symbols[a]	S_S	S_O	X	
Units	C-mmol/L	mmol/L	C-mmol/L	C-mmol X/L h
Process, $j\downarrow$				
1. Growth	$-1/Y_{SX}$	$-1/Y_{OX}$	1	$\mu_{max} = \frac{S_S}{S_S+K_S}X$
2. Decay	0	$-1/\gamma_X$	-1	$k_d X$
Conservation matrix, $k\downarrow$				
1. Carbon (C-mmol)	1	0	1	
2. Nitrogen (N-mmol)	0	0	0.2	
3. Phosphorus (P-mmol)	0	0	0.015	
4. Degree of reduction (mmol e^-/C-mmol)	4	-4	4.13	
5. Charge balance (mmol)	0	0	0	

The composition matrix is explained in the text.
[a]S_S is the substrate concentration, S_O is the oxygen concentration, and X is the biomass concentration.

It is useful to note that each process may act in three different ways on the system components: neutral (zero), negative (consumption) or positive (production). The stoichiometric parameters of the model are defined below in the nomenclature.

The overall component conversion rates, **r**, can then be calculated as the sum of conversion rates of components involved in each process, for example, r for the biomass growth reads as follows:

$$r_3 = r_X = \sum_{j=1}^{n} \mathbf{S}_{3\times j} \cdot \boldsymbol{\rho}_j \tag{4}$$

$$r_X = \mu_{max} \frac{S_S}{S_S + K_S} X - k_d X \tag{5}$$

Finally, the calculated overall component conversion rates can be coupled with appropriate mass balance equations to obtain the ODE for this component. For example, for biomass growth in batch reactors (hence Q is equal to zero), this would be:

$$\frac{dX}{dt} = \mu_{max} \frac{S_S}{S_S - K_S} X - k_d X \tag{6}$$

In a similar fashion, the ODE can be derived for the other system components, that is, substrate and oxygen.

Conservation Principles

In this step, thermodynamic conservation principles—particularly elemental mass and energy balances—are applied to each process of the system as proposed by Roels and coworkers (Roels, 1980; Esener et al., 1983). This provides macroscopic information, which is useful to ensure that there are no errors in the elemental flows in the system and to derive stoichiometric relationships between process variables. Application of such stoichiometric relationships decreases the total number of stoichiometric parameters otherwise needed to describe the process. This aspect is illustrated below (Esener et al., 1983; Heijnen, 1999). A similar method has also been developed successfully by Gujer and Larsen (1995). This conservation principle can be expressed in a matrix form using a composition matrix, **c**:

$$\mathbf{S}_{n\times m} \cdot \mathbf{c}'_{k\times m} = \mathbf{0}_{n\times k} \tag{7}$$

where **c** is the composition matrix and k stands for the total number of conservative properties in the system components. In the example (Table I), the conservative properties are carbon, oxygen, nitrogen, degree of reduction and charge. The composition matrix of the simple aerobic growth model is given at the bottom of Table I. For illustrative purposes, substrate is assumed to be glucose ($C_6H_{12}O_6$) and biomass is assumed to have a general cellular composition of $CH_{1.8}O_{0.5}N_{0.2}P_{0.015}$ (Heijnen, 1999; Roels, 1980). Applying the conservation principles on the simple Monod–Herbert model, it becomes clear that the nitrogen and phosphorus balances are not closed in the model. Of course, if one is not particularly interested in the fate of nitrogen and phosphorus in the system this is no problem. However, in case phosphorus and/or nitrogen are considered explicitly as model components, then one has to ensure that the elemental balance is conserved.

Here we illustrate the application of a conservative property on a process using the matrix notation. As an example, we focus on setting up a macroscopic balance for the degree of reduction ($k = 4$ in the conservation matrix of Table I) for the growth process ($j = 1$ in Table I). Recalling Equation (7), the degree of reduction balance is set-up as follows:

$$\begin{aligned} &\sum_{m=1}^{3} S_{1\times m} c_{4\times m} = 0 \\ &S_{1\times 1} c_{4\times 1} + S_{1\times 2} c_{4\times 2} + S_{1\times 3} c_{4\times 3} = 0 \end{aligned} \tag{8}$$

$$-\frac{1}{Y_{SX}} \gamma_S - \frac{1}{Y_{OX}} \gamma_O + 1\gamma_X = 0 \tag{9}$$

Rearranging Equation (9) and inserting the values of the degrees of reduction gives the following relationship:

$$-\frac{1}{Y_{OX}} = \frac{\gamma_S}{\gamma_O} \frac{1}{Y_{SX}} - \frac{\gamma_X}{\gamma_O} = \frac{4.13}{4} - \frac{1}{Y_{SX}} \tag{10}$$

In the above equations, γ_S, γ_O, and γ_X represent the degree of reduction for substrate, oxygen, and biomass respectively. From the degree of reduction balance, the stoichiometric coefficient of oxygen, Y_{OX}, is estimated using the degrees of reduction and the biomass yield (see Eq. 10). In this way one eliminates the need for the definition of a new parameter, namely Y_{OX}.

Model Development

The model for describing the antibiotic production with *S. coelicolor* was developed and presented using the matrix notation and conservation principles introduced above. While doing so, a special emphasis was given to the creation of a generic and user-friendly framework of model structure development, such that the procedure can be easily repeated and adapted to describe other fermentation processes, for example, different cultures, substrates, operational conditions (pH, temperature), etc.

Stoichiometry and Kinetics of the Biological Processes

Process Stoichiometry

An unstructured approach was employed to describe biological activities on a macro level during the fermentation

process, namely growth, product formation, maintenance and decay processes. Since important cellular components (DNA, RNA, proteins, and enzymes, etc.) are assumed to be in a pseudo-steady state, the growth metabolism can be simplified to:

$$\frac{1}{Y_{SX}}\mathrm{CH}_y\mathrm{O}_z\ (\text{C-mmol}) + z_1\mathrm{O}_2\ (\text{O-mmol}) + z_2\mathrm{NH}_3\ (\text{N-mmol}) + z_3\mathrm{H}_2\mathrm{PO}_4^-\ (\text{P-mmol}) \rightarrow 1\mathrm{CH}_a\mathrm{O}_b\mathrm{N}_c\mathrm{P}_d\ (\text{C-mmol}) + z_4\mathrm{CO}_2\ (\text{C-mmol}) + z_5\mathrm{H}^+\ (\text{H-mmol}) + (\ldots)\mathrm{H}_2\mathrm{O} \quad (11)$$

The above biochemical reaction is written based on production of 1 unit of biomass, and is related to substrate consumption using the biomass growth yield on substrate, Y_{SX}, which was experimentally determined. Moreover, in this study, ammonia was assumed as nitrogen source for the biomass growth (Heijnen, 1999). Stoichiometric coefficients of all other components (z_1, ..., z_5) were determined using the conservation principles of the degree of reduction, nitrogen, phosphorus, carbon and charge respectively (Heijnen, 1999). The complete list of stoichiometric coefficients for the biomass growth process is shown in Table IIA (process number 1).

In a similar fashion, product formation metabolism at pseudo-steady state would be:

$$\frac{1}{Y_{SP}}\mathrm{CH}_y\mathrm{O}_z\ (\text{C-mmol}) + y_1\mathrm{O}_2\ (\text{O-mmol}) + y_2\mathrm{NH}_3\ (\text{N-mmol}) \rightarrow 1\mathrm{CH}_p\mathrm{O}_q\mathrm{N}_v\ (\text{C-mmol}) + y_3\mathrm{CO}_2\ (\text{C-mmol}) + y_4\mathrm{H}^+\ (\text{H-mmol}) + (\ldots)\mathrm{H}_2\mathrm{O} \quad (12)$$

The substrate consumption is related to product formation through the experimentally determined product yield on substrate, Y_{SP}. All other stoichiometric coefficients were determined from conservation principles (see above).

In order to have a comprehensive description of the maintenance during batch fermentations, the Herbert and Pirt models were both employed in a complimentary way (see Results and Discussion Section). In the Pirt maintenance model (Pirt, 1965), external substrate is consumed to generate sufficient energy required for the maintenance of biomass activities:

$$1\mathrm{CH}_y\mathrm{O}_z\ (\text{C-mmol}) + u_1\mathrm{O}_2\ (\text{O-mmol}) \rightarrow 1\mathrm{CO}_2\ (\text{C-mmol}) + (\ldots)\mathrm{H}_2\mathrm{O} \quad (13)$$

In the Herbert model, the maintenance energy is supplied by oxidizing part of the biomass (internal macromolecules or polymers) leading to a decrease in the total biomass in the system, that is, a negative growth (Esener et al., 1983):

$$1\mathrm{CH}_a\mathrm{O}_b\mathrm{N}_c\mathrm{P}_d\ (\text{C-mmol}) + w_1\mathrm{O}_2\ (\text{O-mmol}) \rightarrow 1\mathrm{CO}_2\ (\text{C-mmol}) + w_2\mathrm{NH}_3\ (\text{N-mmol}) + w_3\mathrm{H}_2\mathrm{PO}_4^-\ (\text{P-mmol}) + (\ldots)\mathrm{H}_2\mathrm{O} \quad (14)$$

In the decay process, nitrogen and phosphorus are released to the external medium along with some inert material upon cell death or lysis (Henze et al., 2000). The experimental observations from the fermentation studies will determine whether or not nitrogen and phosphorus release should be taken into account when applying the Herbert biomass decay model. In our model we excluded the release of nitrogen and phosphorus by setting the coefficients w_2 and w_3 to zero, based on the experimental observations (see below). Nonetheless, to have a generally applicable model structure the above mentioned decay stoichiometry was retained in the model matrix. The overall stoichiometric matrix of the process model is summarized in Table IIA.

Table IIA. Stoichiometric matrix of the biological process model*.

Components → i	1	2	3	4	5	6	7	8	9
Name	Glucose	Oxygen	Ammonia	Phosphate	Biomass (*S. coelicolor*)	Antibiotic 1 (ACT)	Antibiotic 2 (RED)	Carbon dioxide	Hydrogen ion
Symbol	S_g	S_O	S_{NH_3}	S_{PO}	X	S_{P1}	S_{P2}	S_{CO_2}	S_H
Chemical composition	$C_6H_{12}O_6$	O_2	NH_3	$H_2PO_4^-$	$CH_{1.8}O_{0.5}N_{0.2}P_{0.015}$	$C_{32}H_{26}O_{14}$	$C_{25}H_{35}N_3O$	CO_2	H^+
	Units								
j Processes↓	C-mmol/L	O-mmol/L	N-mmol/L	P-mmol/L	C-mmol/L	C-mmol/L	C-mmol/L	C-mmol/L	H-mmol/L
1 Biomass growth	$-1/Y_{SX}$	$\gamma_X/4 - \gamma_S/(Y_{SX}\times 4)$	$-i_{NX}$	$-i_{PX}$	1			$1/Y_{SX} - 1$	$-i_{PX}$
2 Actinorhodin production	$-1/Y_{SACT}$	$\gamma_{ACT}/4 - \gamma_S/(Y_{SACT}\times 4)$				1		$1/Y_{SACT} - 1$	
3 Undecylprodigiosin production	$-1/Y_{SRED}$	$\gamma_{RED}/4 - \gamma_S/(Y_{SRED}\times 4)$	$-i_{NRED}$				1	$1/Y_{SRED} - 1$	
4 Biomass maintenance	-1	$-\gamma_S/4.0$						1	
5 Biomass decay		$-\gamma_X/4$	i_{NX}	i_{PX}	-1			1	i_{PX}

*All the empty cells of this matrix should be read zero, while positive or negative coefficients should be read as production or consumption, respectively.

Process Rates

The kinetics of the above-mentioned biological processes have been described in the literature using simple unstructured models. The biomass growth kinetics was described using the Monod model with the assumption of multiple substrate limitations—typically done in this field. In batch processes the assumption of balanced growth is not valid particularly during transient conditions, that is, the lag-phase which precedes the exponential growth phase. While several structured, yet complicated models have been developed to describe this lag-phase (Daae and Ison, 1998; King, 1997; Mundry and Kuhn, 1991; among others), we chose to use an empirical yet simple expression based on a sigmoidal type of function (a switch function) which requires only one parameter, that is, $1/(1+e^{t_{lag}-t})$ (see process rate for the biomass growth ($j=1$) in Table IIB). In this expression, t stands for time and the parameter t_{lag} represents the onset time of the exponential growth phase.

The substrate consumption and the CO_2 production kinetics during the growth process were described using linear relations based on the growth kinetics. This approach assumes that the growth kinetics is the rate limiting step in the overall metabolism (Noorman et al., 1991).

Product formation kinetics was described using a modified version of the Luedeking–Piret model. The non-growth associated product formation part of the Luedeking–Piret model was changed to a logistic type expression (see Table IIB, processes 2 and 3). This choice was motivated by the experimental observations collected in this study and the results of Kiviharju et al. (2006). The onset of product formation was mathematically formulated using a switch function based on the inverted Monod expression of phosphate limitation, that is, $K_{IP}/(K_{IP}+S_{PO})$ (Table IIB, processes 2 and 3). This particular expression also describes the decrease in the specific growth of biomass under phosphate limited growth conditions, which is currently believed to be an important triggering mechanism behind secondary metabolism (Doull and Vining, 1990). Phosphate limitation is also thought to be directly involved in the regulation of the secondary metabolism of *S. coelicolor* (Martín, 2004). Hence this particular expression of phosphate limitation reflects the above-mentioned mechanistic hypothesis about the onset of antibiotic production.

The Pirt maintenance kinetics was modified with the addition of two Monod substrate limitation functions for oxygen and substrate (see Table IIB, process 4). The first order kinetics of the Herbert model was modified with the addition of electron acceptor limitation and a substrate concentration dependent switch function, that is, $K_S/(K_S+S_S)$ (see Table IIB, process 5). The motivation and the underlying hypothesis for these modifications are discussed in Biomass Maintenance in Batch Fermentations: Pirt Versus Herbert Section. The overall kinetic rates of the biological processes are summarized in Table IIB.

Chemical Processes: Kinetic Description of Mixed Weak Acid–Base Systems

The pH effect of biological processes has been well studied with the aim of exploiting the pH effect for monitoring, control and understanding of fermentation processes (Royce, 1992; San and Stephanopoulos, 1984; Siano, 1995). Building on these earlier studies which mainly focused on fermentations operated with a fixed pH set-point (pH-stat) and equilibrium-chemistry, we employed a mixed weak acid/base kinetic model. This approach was chosen since it enables dynamic prediction of pH during the fermentation process. To this end, we make use of the mixed weak acid/base kinetic model of Musvoto et al. (1997), which has been successfully applied for dynamic pH prediction in biological wastewater treatment plants (Sötemann et al., 2005).

In this chemical model, all relevant weak acid/base species, that is, those species that are consumed or produced in a significant quantity—hence significantly affect the pH—during the fermentation process are considered. For typical fermentation processes these species would be ammonium, phosphate, bicarbonate and short-chain fatty acids, for example, acetate. The dissociation kinetics of these weak acid/bases is described by summing up forward and reverse reaction rates. For example, the dissociation of ammonium/ammonia is given by

$$NH_4^+ \leftrightarrow NH_3 + H^+ \tag{15}$$

Table IIB. The rate vector of the biological process model.

j	Processes	Rates (mmol/L day)
1	Biomass growth	$\mu_{max}\frac{1}{1+e^{t_{lag}-t}}\frac{S_g}{S_g+K_S}\frac{S_O}{S_O+K_O}\frac{S_{NH_3}}{S_{NH_3}+K_{NH_3}}\frac{S_{PO}}{S_{PO}+K_{PO}}X$
2	Actinorhodin production	$\alpha_{ACT}r_X+\beta_{ACT}\left(1-\frac{S_{ACT}}{S_{ACT}^{max}}\right)\left(\frac{S_g}{K_S+S_g}\frac{K_{IP}}{K_{IP}+S_{PO}}\right)$
3	Undecylprodigiosin production	$\alpha_{RED}r_X+\beta_{RED}\left(1-\frac{S_{RED}}{S_{RED}^{max}}\right)\left(\frac{S_g}{K_S+S_g}\frac{K_{IP}}{K_{IP}+S_{PO}}\right)X$
4	Biomass maintenance	$m_S\frac{S_g}{S_g+K_S}\frac{S_O}{S_O+K_O}X$
5	Biomass decay	$k_d\frac{S_O}{S_O+K_O}\frac{K_S}{S_g+K_S}X$

The overall kinetics of ammonium dissociation is given by the sum of forward and reverse reactions (Musvoto et al., 1997):

$$r_{NH} = k_{f,NH} S_{NH_4} - \frac{k_{f,NH}}{K_{NH}} S_{NH_3} S_H \quad (16)$$

In a similar fashion, the dissociation kinetics for the carbon dioxide buffer, phosphate buffer and water were derived. For the buffer systems, those dominant species at pH range 5–8 (i.e., CO_2–HCO_3^- for the carbon dioxide buffer system and $H_2PO_4^-$–HPO_4^{2-} for the phosphate buffer system) were considered in this study, since this range is relevant for most operating conditions in fermentation applications. The stoichiometry and the rate expressions of the mixed weak acid/base kinetic model are provided in Table III.

Physical Model: Gas–Liquid Exchange Processes

The physical part of the integrated model deals with the gas–liquid exchange processes. Of major concern for fermentation processes are the mass transfer of oxygen from gas to liquid and the stripping of dissolved CO_2 from liquid to gas phase. For example, the gas–liquid exchange process for CO_2 reads as follows:

$$CO_2(\text{gas}) \leftrightarrow CO_2(\text{liquid}) \quad (17)$$

The mass transfer rate of CO_2 from gas to liquid (or vice versa) can be modeled according to the two-film theory, which states that the transfer is driven by a gradient between the saturation concentration of CO_2 in the liquid phase (equilibrium) and the actual CO_2 concentration in the liquid:

$$r_{CO_2 \to L} = K_L a_{CO_2} (S^*_{CO_2} - S_{CO_2}) \quad (18)$$

The equilibrium (or saturation) concentration of CO_2 in the liquid is proportional to the partial pressure of CO_2 in the gas phase (Henry's law) and given by

$$S^*_{CO_2} = P_{CO_2} K_{HCO_2} \quad (19)$$

In a similar fashion, the mass transfer rates for oxygen, nitrogen and ammonia were derived and provided in Table IV. The stoichiometric matrix of the gas–liquid exchange processes is also shown in Table IV.

Summarizing Process Stoichiometry and Kinetics of the Integrated Biological, Chemical and Physical Model

The different sub-models presented above were subsequently coupled together in one single matrix to form an integrated biological, chemical and physical model (available for download as supplementary material). The conservation matrix was also included in this complete matrix of the model. The integrated model contains in total around 60 parameters, which need to be provided with nominal values to enable successful simulation of the fermentation process.

Parameters of the Model

The parameters of the integrated model are shown in Table V together with the corresponding literature sources from which they were obtained. The stoichiometric parameters were either estimated from the experimental data (see below) or directly calculated from the composition of substrates, products and biomass, for example, molecular weight, degree of reduction, nitrogen content, phosphorus content and the like. The product yields were taken from the literature (see, e.g., Bruheim et al., 2002).

For the kinetic parameters of the model, the endogenous decay coefficient was estimated directly from the dry-weight measurements of biomass during the death phase of the typical fermentation process (available as Supplementary Material). The maximum growth and antibiotic production rates and the affinity coefficient of phosphate were estimated

Table III. The stoichiometric and kinetic matrix of the mixed weak acid/base model.

Components → i	3	4	8	9	10	11	12	13	
Name	Ammonia	Phosphate	Carbon dioxide	Hydrogen	Ammonium	Phosphate	Bicarbonate	Hydroxyl ion	Process rates
Symbol	S_{NH}	$S_{H_2PO_4}$	S_{CO_2}	S_H	S_{NH_4}	S_{HPO_4}	S_{HCO_3}	S_{OH}	
Chemical composition	NH_3	$H_2PO_4^-$	CO_2	H^+	NH_4^+	HPO_4^{2-}	HCO_3^-	OH^-	
					Units				
j Processes↓	N-mmol/L	P-mmol/L	C-mmol/L	H-mmol/L	N-mmol/L	P-mmol/L	C-mmol/L	H-mmol/L	mmol/L day
6 Ammonium dissociation	1			1	−1				$k_{f,NH} S_{NH_4} - \frac{k_{f,NH}}{k_{NH}} S_{NH_3} S_H$
7 Phosphate dissociation		−1		1		1			$k_{f,H_2PO_4} S_{H_2PO_4} - \frac{k_{f,H_2PO_4}}{k_{H_2PO_4}} S_{HPO_4} S_H$
8 Carbon dioxide dissociation			−1	1			1		$k_{f,CO_2} S_{CO_2} - \frac{k_{f,CO_2}}{k_{HCO_3}} S_{HCO_3} S_H$
9 Water dissociation				1				1	$1 - \frac{k_{f,w}}{K_W} S_H S_{OH}$

Table IV. The stoichiometric and kinetic matrix of the physical model: gas–liquid exchange processes.

Components → i	2	3	8	14	15	16	17	18	
Name	Oxygen dissolved	Ammonia dissolved	Carbon dioxide dissolved	Nitrogen dissolved	Oxygen gas	Carbon dioxide gas	Nitrogen gas	Ammonia gas	Process rates
Symbol[a]	S_O	S_{NH_3}	S_{CO_2}	S_{N_2}	G_O	G_{CO_2}	G_{N_2}	G_{NH_3}	
Chemical composition	O_2	NH_3	CO_2	N_2	O_2	CO_2	N_2	NH_3	
					Units				
j Processes↓	O-mmol /L	N-mmol /L	C-mmol /L	N-mmol /L	O-mmol /L	C-mmol /L	N-mmol /L	N-mmol /L	mmol/L day
10 Aeration (oxygen)	1				−1				$K_La_{O_2}(S^*_O - S_O)$
11 CO_2 stripping			1			−1			$K_La_{CO_2}(S^*_{CO_2} - S_{CO_2})$
12 Nitrogen stripping				1			−1		$K_La_{N_2}(S^*_{N_2} - S_{N_2})$
13 Ammonia stripping		1						−1	$K_La_{NH_3}(S^*_{NH_3} - S_{NH_3})$

[a]The gas phase components are denoted by a capital letter G with subscripts indicating the gaseous species.

from the data (see below). The rest of the kinetic parameters, that is, the half-saturation coefficients were taken from literature (see Table V). The latter is due to the fact that these parameters are not identifiable under substrate unlimited conditions, unless a dedicated design of experiments is performed (Baltes et al., 1994; Holmberg, 1982; Munack, 1991). However, such a design of experiments is beyond the scope of this study.

Since temperature dependency is important for most of the chemical reaction kinetics, parameter values were given as a function of temperature. The apparent equilibrium constants of chemical reactions and their temperature dependencies were adopted from reference chemical databases. Furthermore, as the kinetics of most chemical equilibrium reactions are achieved much faster compared to the time constant of the biological processes, the apparent

Table V. Parameters of the integrated model of *S. coelicolor* fermentation (see the units in the nomenclature; values in italics are estimated).

Symbol	Value	References	Symbol	Value	References
Y_{SX}	0.5	*Parameter estimation*	T	303.15	Measurement
i_{NX}	0.22	*Parameter estimation*	pH	7.02	Measurement
i_{PX}	0.015	*Parameter estimation*	pK_{NH}	$2835.8/T - 0.6322 + 0.00123T$	Musvoto et al. (1997)
MW_X	25.065	Biomass composition	pK_{HCO_3}	$3404.7/T - 14.8435 + 0.03279T$	Musvoto et al. (1997)
MW_S	30.0	Chemical composition of glucose	$pK_{H_2PO_4}$	$1979.5/T - 5.3541 + 0.01984T$	Musvoto et al. (1997)
γ_X	4.125	Biomass composition	pK_w	14.0000	Musvoto et al. (1997)
γ_S	4.0	Composition of glucose	k_{fbNH}	1.0E+07	Musvoto et al. (1997)
f_{XI}	0.09	Biomass composition	k_{f,CO_2}	1.0E+07	Musvoto et al. (1997)
μ_{max}	0.21	*Parameter estimation*	k_{f,H_2PO_4}	1.0E+07	Musvoto et al. (1997)
m_S	0.041	Elibol and Mavituna (1999)	k_{fbW}	1.0E+07	Musvoto et al. (1997)
k_d	0.02	Prolonged batch tests (Supplementary 1)	R	0.082057	Universal gas constant
t_{lag}	39.0	*Parameter estimation*	D_{O_2}	2.42E−05	CRC Handbook (2006–2007)
K_S	0.033	Presumed not rate limiting	D_{CO_2}	1.91E−05	CRC Handbook (2006–2007)
K_O	0.031	Presumed not rate limiting	D_{N_2}	2.00E−05	CRC Handbook (2006–2007)
K_{NH_3}	0.001	Presumed not rate limiting	P_{in}	*1.22*	*Parameter estimation*
K_{PO}	1.9	*Parameter estimation*	P_{O_2}	0.2098	Measurement
Y_{SACT}	0.125	Bruheim et al. (2002)	P_{CO_2}	0.0003	Measurement
α_{ACT}	0.0	Experimental observations	P_{NH_3}	0.0	Sötemann et al. (2005)
β_{ACT}	3.2E−3	*Parameter estimation*	P_{N_2}	$(1 - P_{O_2} - P_{CO_2} - P_{NH_3})$	CRC Handbook (2006–2007)
S^{max}_{ACT}	5.0480	Experimental observations	K_{HO_2}	$0.0013\, e^{(-1{,}700\times(1/T-1/298.15))}$	NIST Chemistry WebBook (2005)
K_{IP}	0.1	Experimental observations	K_{HCO_2}	$0.045\, e^{(-2{,}400\times(1/T-1/298.15))}$	NIST Chemistry WebBook (2005)
MW_{act}	19.81	Chemical composition of ACT	K_{HN_2}	$0.00065\, e^{(-1{,}700\times(1/T-1/298.15))}$	NIST Chemistry WebBook (2005)
γ_{ACT}	3.94	Chemical composition of ACT	$S^*_{O_2}$	$P_{O_2}K_{HO_2}$	Henry's law
Y_{SRED}	0.0954	Assumed	$S^*_{CO_2}$	$P_{CO_2}K_{HCO_2}$	Henry's law
α_{RED}	0.0	Experimental observations	$S^*_{N_2}$	$P_{N_2}K_{NH_2}$	Henry's law
β_{RED}	4.8E−5	*Parameter estimation*	$S^*_{NH_3}$	0	Sötemann et al. (2005)
S^{max}_{RED}	0.12	Experimental observations	$K_La_{O_2}$	*137*	*Parameter estimation*
MW_{red}	15.72	Chemical composition of RED	$K_La_{CO_2}$	$K_La_{O_2}\left(\frac{D_{CO_2}}{D_{O_2}}\right)^{0.5}$	Empirical relationship
γ_{RED}	4.96	Chemical composition of RED	$K_La_{N_2}$	$K_La_{O_2}\left(\frac{D_{N_2}}{D_{O_2}}\right)^{0.5}$	Empirical relationship
i_{NRED}	0.12	Chemical composition of RED	$K_La_{NH_3}$	3	Assumption (Sötemann et al., 2005)

forward reaction rates of chemical equilibrium were set high as suggested by Musvoto et al. (1997).

The mass transfer coefficient of CO_2 and nitrogen gases is commonly estimated using an empirical relationship that uses oxygen as the reference gas:

$$K_L a_{CO_2} = K_L a_{O_2} \left(\frac{D_{CO_2}}{D_{O_2}} \right)^n \tag{20}$$

The Henry's coefficient and the molecular diffusion coefficients of gases in water were obtained from standard chemical and physical reference databases (CRC Handbook, 2006–2007; NIST Chemistry WebBook, 2005).

Liquid and Gas Phase Mass Balances

The general mass balance of model components in the liquid phase was obtained straightforwardly using Equation (2). The overall component conversion rates vector, **r**, was obtained from the model matrix (see Eq. 1). This resulted in a set of p ODEs, that is, one for each model component in the liquid phase.

The rate of change of gaseous components (total number equal to $m-p$) was determined from the gas-phase mass balances. To this end, the gas-phase was assumed to be completely mixed. Hence, the mass balances for the gaseous components can be written as follows:

$$\frac{d(C_{G,j}V_G)}{dt} = Q_G^{in}C_{G,j}^{in} - Q_G^{out}C_{G,j} - r_{G\to L,j}V_L \quad \text{for } j = p+1, p+2, \ldots, m \tag{21}$$

In Equation (21) the subscript $p+1$ indicates the indexes of the gaseous components in the overall model, and m is the total number of model components and other parameters are as explained in the nomenclature. The molar concentration of the constituents of the air flow-rate was calculated using the ideal gas law. This can be done given the incoming air pressure, P_G^{in} and temperature, T_G^{in} are known:

$$C_G^{in} = \frac{n_G}{V_G} = \frac{P_G^{in}}{RT_G^{in}} \tag{22}$$

In the off-gas stream, the partial pressure of each gaseous component in the mixture was calculated from Dalton's partial pressure law as follows:

$$P_{G,j} = P_G^{out}C_{G,j} \tag{23}$$

Finally the off-gas flow rate, Q_G^{out}, was calculated from the overall gas-phase mass balance around the fermenter, which reads as (Volcke, 2006):

$$\frac{dn_G}{dt} = n_G^{in} - n_G^{out} - n_G^{exchange} \tag{24}$$

Assuming that the gas-phase is at steady state, the above equation can then be simplified to the following expression:

$$\begin{aligned} n_G^{out} &= n_G^{in} - n_G^{exchange} \\ Q_G^{out}\frac{P_G^{out}}{RT_G^{out}} &= Q_G^{in}\frac{P_G^{in}}{RT_{in}} - \sum_{j=p+1}^{m} r_{G\to L,j}V_L \end{aligned} \tag{25}$$

In this study, the partial pressure of water vapor in air was excluded and the total pressure of the off-gas mixture was equal to atmospheric pressure. The off-gas temperature was assumed equal to the liquid temperature in the fermenter, while the remaining parameters of the equation are as explained above. The overall gas mass balance (Eq. 25) was numerically solved (iteratively) in combination with the mass balances for the individual gas components (Eq. 21).

Materials and Methods

Strain and Inoculum

The reference strain was *S. coelicolor* M145 ($SCP1^-$, $SCP2^-$), a prototrophic non-plasmid containing derivative of *S. coelicolor* A3 (2) strain 1147 (Bentley et al., 2002; Redenbach et al., 1996). The strain was a kind gift from Mervyn Bibb (John Innes Centre, Norwich, UK).

For preparation of inoculum, *S. coelicolor* spores were densely plated on mannitol soya flour medium (Kieser et al., 2000). The spores from each plate were harvested by addition of 1 mL 20% glycerol and used to inoculate 50 mL of 2xYT medium (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl) in 200 mL shake flasks with 20 glass beads of 0.3 mm diameter. The cultures were grown at 30°C on an orbital shaker with agitation rate 150 rpm for 24 h. The mycelium was collected by centrifugation at 4,000*g* for 5 min and the supernatant decanted. The mycelium was mixed with the residual liquid and crushed in a crusher, consisting of a glass cylinder and a tight glass plunger. Crashed mycelium was resuspended in 20% peptone to an optical density of 26 (600 nm) and stored at −20°C until further use as inoculum.

Batch Cultivations

The cultivations were carried out in a 3.5 L stainless steel fermenter (MBR Bioreactor, Zürich, Switzerland) with a working volume of 3 L and operated with 1 vvm aeration and a stirring rate of 600 rpm (one 6-bladed Rushton turbine). The temperature was kept at 30°C, and pH (465-50-SC-P-S7 sensor from Mettler Toledo International Inc., Greifensee, Switzerland) was controlled at 7 by addition of 2 M NaOH. The dissolved oxygen tension in the solution was monitored with an oxygen electrode (Mettler Toledo), and the partial pressure of the off-gas CO_2 was monitored using a gas analyzer (Brüel & Kjaer, Naerum, Denmark). The

cultivation medium contained 2 mM Na_2SO_4, 2 mM Na-citrate, 10 mM KCl, 2 mM $MgCl_2$, 1.25 mM $CaCl_2$ and 3 mM NaH_2PO_4. In addition, 115 mM NH_4Cl and 30 g/L glucose were added to the medium in batch 1, while 100 mM NH_4Cl and 40 g/L glucose were added to the medium in batch 2. The basic salt solution was supplemented with 15 mL/L trace metal solution containing 20 mM $FeCl_3$, 10 mM $CuCl_2$, 50 mM $ZnCl_3$, 10 mM $MnCl_2$, 0.02 mM Na_2MoO_4, 20 mM $CoCl_2$, 10 mM H_3BO_4, as well as with 1 mL/L vitamin solution (Verduyn et al., 1992). The fermenter was inoculated with 10 mL mycelium suspension prepared as described above, which had been stored at $-20^{\circ}C$ and was thawed just before inoculation.

Off-Line Analysis

For cell dry weight measurements, a known amount of sample was filtered through a pre-weighted 0.45 μm pore size filter (Pall Corporation, Ann Arbor, MI), which was washed with an equal volume of 0.9% NaCl and dried for 20 min in a microwave oven at 150 W before the weight gain was recorded. Glucose was analyzed using a HPLC system (Dionex Softron GmbH, Germenring, Germany) equipped with an Aminex HPX-87H column (Bio-Rad Laboratories, Hercules, CA) operating at 60°C. The compounds were eluted with 5 mM H_2SO_4 at a flow rate of 0.6 mL/min and glucose was quantified using a differential refractometer (Shimadzu, Kyoto, Japan). Ammonium was determined using an ammonia-selective electrode (Metrohm, Herisau, Switzerland) and the phosphate concentration was determined photometrically by measuring the absorbance of the malachite green complex formed in the presence of molybdate using a Pi ColorLock Gold kit (Innova Biosciences, Cambridge, UK). For determination of the antibiotics, spectrophotometric based methods were applied. The blue actinorhodin pigment was extracted by vortexing the sample with 3 N KOH, the total volume of the mixture being 1.5 mL. The sample volume was from 0.75 to 0.075 mL, decreasing for the higher actinorhodin concentrations in the broth. The mixture was centrifuged at 10,000*g* for 5 min and the absorbance of the supernatant was measured at 640 nm. The concentration of actinorhodin was calculated using an extinction coefficient of 25,320 $M^{-1}\,cm^{-1}$. For measuring the red prodigiosin pigments, 0.75 mL of the broth was vortexed with an equal volume of methanol buffered to pH 1.5 with concentrated hydrogen chloride. The extraction was carried out under shaking for 3 h. The mixture was centrifuged as above and the absorbance of the supernatant was measured at 530 nm. The concentration was found with an extinction coefficient of 100,500 $M^{-1}\,cm^{-1}$.

Modeling and Simulation

The model has been implemented in Matlab (The Mathworks, Natick, MA). The ODEs were solved using a stiff-solver (ode15s with integration accuracy set to 1.0E−07).

Parameter Estimation

A non-linear least squares method was used for the parameter estimation. The objective function used for the parameter estimation was constructed as the weighted sum of squared errors between the measurements and their corresponding model predictions (Seber and Wild, 1989):

$$J(\boldsymbol{\theta}) = \sum_{k=1}^{N}\sum_{t=1}^{M}\left(\frac{y_{k,\mathrm{meas}}(t) - y_k(t,\boldsymbol{\theta})}{\sigma_{k,t}}\right)^2 \tag{26}$$

In this study, the standard deviation of measurements, $\sigma_{k,t}$, was assumed to be 10% of the average measured concentration during the batch. The standard deviation of measurements was assumed to be identical at each time instant. The objective function was minimized by using the trust-region based nonlinear minimization algorithm in Matlab.

Confidence Interval, Correlation, and Uncertainty

A linear approximation of the covariance matrix of parameter estimators, $\mathbf{COV}(\boldsymbol{\theta})$, was used to estimate the confidence interval of the estimated parameters (Seber and Wild, 1989). The covariance matrix of the estimated parameters is approximated as (Omlin and Reichert, 1999):

$$\mathbf{COV}(\boldsymbol{\theta}) = \frac{J(\boldsymbol{\theta})}{NM-Q}\left(\left(\frac{\partial \mathbf{y_k}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}}\begin{pmatrix}\frac{1}{\sigma_{1,1}^2} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \frac{1}{\sigma_{k,K\times T}^2}\end{pmatrix}\left(\frac{\partial \mathbf{y_k}}{\partial \boldsymbol{\theta}}\right)\right) \tag{27}$$

The confidence interval of the estimated parameters at $(1-\alpha)$ confidence level is given by

$$\boldsymbol{\theta}_{1-\alpha} = \boldsymbol{\theta} \pm \sqrt{\mathrm{diag}(\mathbf{COV}(\boldsymbol{\theta}))}\cdot t\left(N-M, \frac{\alpha}{2}\right) \tag{28}$$

where $t(N-M, \alpha/2)$ is the t-distribution value corresponding to the $\alpha/2$ percentile with $N-M$ degrees of freedom, and diag represents the diagonal elements of the covariance matrix of the parameters.

The linear correlation between two parameters, R_{ij}, is given by

$$R_{ij} = \frac{\mathrm{COV}(\theta_i,\theta_j)}{\sqrt{\sigma_{\theta_i}^2\sigma_{\theta j}^2}} \tag{29}$$

The confidence interval of the predictions is calculated using the covariance matrix of estimated parameters. First, the covariance matrix of predictions is approximated using

linear error propagation as follows (Omlin and Reichert, 1999):

$$\mathbf{COV}(\mathbf{y_k}) = \left(\frac{\partial \mathbf{y_k}}{\partial \boldsymbol{\theta}}\right) \cdot COV(\theta) \cdot \left(\frac{\partial \mathbf{y_k}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}} \qquad (30)$$

Then the confidence interval of the predictions at $(1-\alpha)$ confidence level is given by

$$\mathbf{y}_{k,1-\alpha} = y_k \pm \sqrt{\mathrm{diag}(\mathbf{COV}(\mathbf{y_k}))} \cdot t\left(N-M, \frac{\alpha}{2}\right) \qquad (31)$$

The above-mentioned methods for the calculation of parameter estimation errors and the correlation matrix were implemented in Matlab as m-file scripts.

Results and Discussion

The integrated model was evaluated in view of dynamic description of the physical-chemical (off-gas CO_2, dissolved oxygen and base addition measurements) and biological process data (glucose, biomass, ammonium, phosphate, actinorhodin and undecylprodigiosin measurements) obtained from two batch cultivations operated under phosphate limitation (see the data in Figs. 1–3). It is important to clarify that while ammonia-nitrogen was assumed to be taken up for biomass growth (see Table IIA), the model fits however were compared to the ammonium-nitrogen measurements as these were available. This is perfectly possible since the dissociation of ammonium into ammonia and proton species was explicitly considered in the model structure and assumed instantaneous (see Table III). Hence ammonium concentration (dynamics) in the liquid phase will follow instantaneously the ammonia uptake dynamics (see Supplementary Material 3), thereby allowing us to compare the model description of nitrogen uptake of biomass growth to the measured ammonium concentrations in the liquid.

Batch 1 was run with 30 g/L glucose as initial substrate concentration, while batch 2 was run with 40 g/L glucose. The biomass concentration obtained in batch 2 was double the amount obtained in batch 1, whereas for the antibiotic an opposite trend was seen, batch 2 produced a lot less than batch 1 (see the data in Figs. 1 and 2). Below we present the results of the model identification and the quality of the model fits and discuss in detail the interpretation of the results in view of enhancing the process understanding as expected from a PAT application.

Model Identification on Typical Fermentation Batch Data

The parameter subset used in the model identification included stoichiometric (i.e., biomass yield, nitrogen and phosphate content), biokinetic (maximum growth rate, Monod affinity constant, antibiotic production rates) and operational parameters (inlet pressure and mass transfer coefficient of oxygen, related to the intensity of aeration) (Table VI). The parameters were estimated simultaneously using the aforementioned measurements of biological, chemical and off-gas data. The resulting parameter estimates, the confidence intervals and the correlation matrix are provided in Table VI.

One observes that some of the parameter estimates remained similar in both batches, including the yield and inlet air pressure. The estimates of the other parameters, however, differed considerably between the two batches (Table VI), which can be explained by three possible reasons: (i) identifiability, (ii) operational, and (iii) initial conditions of the batch cultivations. The differences between both batches, particularly, between the estimates of the maximum growth rate and the Monod affinity constant for phosphate are attributed to an identifiability issue (see below). This is indicated by the high correlation coefficient found between the estimates of the maximum growth rate and the Monod affinity coefficient (see Table VI, R was 0.99), which resulted in a large 95% confidence interval (almost $\pm$100% of the mean estimated value). In short, while the mean estimated values of $\mu_{\max}$ and K_{PO} were different in both batches, these mean values however were within their 95% confidence range (compare the mean value of $\mu_{\max}$ in batch 1 to the 95% confidence range of $\mu_{\max}$ in batch 2, Table VI). As for the difference between the estimates of the mass transfer coefficient of oxygen, this can simply be caused by an operational difference between the two batches, for example, a decrease in air flow rate. On the other hand, the differences between the estimates of the remaining parameters (especially the lag-time for the onset of biomass growth, the nitrogen content, phosphate content and the actinorhodin production rate) are believed to be caused by the differences in the initial conditions of both batch cultivations of *S. coelicolor*.

The confidence interval indicates the quality of parameter estimates, where a smaller interval means a high quality of the parameter estimate or equally a small error. The quality of parameter estimates obtained from both batches was comparable for most parameters with the exception of the maximum growth rate, the affinity constant of phosphate (already discussed above) and the actinorhodin production rate coefficient. While most of the estimated parameters had a relative error of about 5–10% of their mean value, the confidence intervals of the maximum growth rate, affinity constant of phosphate and actinorhodin production rate were larger (e.g., the confidence interval of the affinity coefficient of phosphate was found to be around 50% and 100% in batch 1 and batch 2, respectively). The quality of parameter estimates depends on the model structure, and the quality and quantity of the data used for identification (Baltes et al., 1994; Holmberg, 1982; Munack, 1991). The quality of parameter estimates can in theory be further improved by considering optimal design of experiments, but this is not within the scope of this contribution.

The correlation matrix of all the estimated parameters is shown in Table VI for batch 2 (the correlation matrix obtained for the estimates using data of batch 1 was similar). The correlation between parameters can vary between

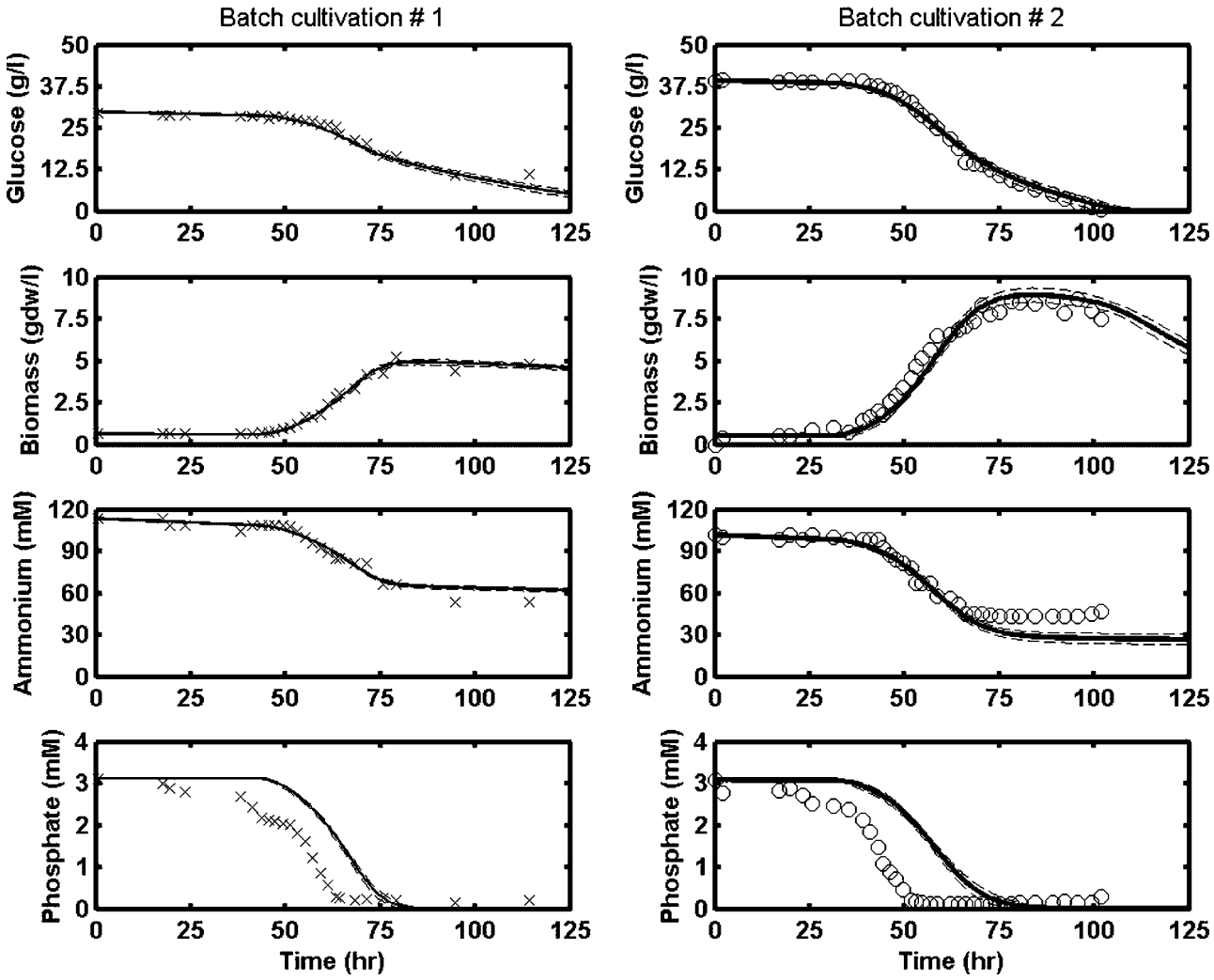


Figure 1. Model fits to the biomass, glucose, ammonium and phosphate measurements together with 95% upper and lower confidence bounds: (×) data batch 1, (–) model batch 1, (○) data batch 2, and (▬) model batch 2.

±1 indicating complete correlation and 0 indicating no correlation. Apart from the high correlation observed between the maximum specific growth rate, μ_{max}, and the affinity constant for phosphate, K_{PO}, one also observes that these two parameters were significantly correlated with the lag time, t_{lag} (around 75% see Table VI). The correlations between the remaining parameters were found to be not that significant (assuming 0.70 as a cut off value for R to be significant). The parameter estimation errors and the correlation matrix were taken into account appropriately by the uncertainty analysis (see below).

Evaluation of the Model Fits

The quality of the model fits is assessed by considering their corresponding accuracy and uncertainty. Generally speaking the accuracy relates to how well the model describes data, while the uncertainty relates to error on the fits. The accuracy of the model fits is evaluated by the mean squared error (MSE) criterion, and values for this criterion are shown in Table VII. The MSE criterion is expected to be low for good model fits (Power, 1993). The uncertainty of the fits is obtained by linear propagation of the parameter estimation errors (see Eqs. 27–31) and also expected to be low. The uncertainty is represented by 95% uncertainty bounds around the fits in Figures 1–3. In these figures, the middle line indicates the average fitted value, while the upper and lower lines indicate the error on the model fits at 95% confidence level (Omlin and Reichert, 1999). The units of biomass, glucose, actinorhodin and undecylprodigiosin were converted from mol basis to mass using their corresponding molecular weights (see Table V).

The Glucose, Ammonium, Phosphate Uptake and Biomass Growth Kinetics

Model fits to the glucose, ammonium, phosphate and biomass data are shown for both batches in Figure 1.

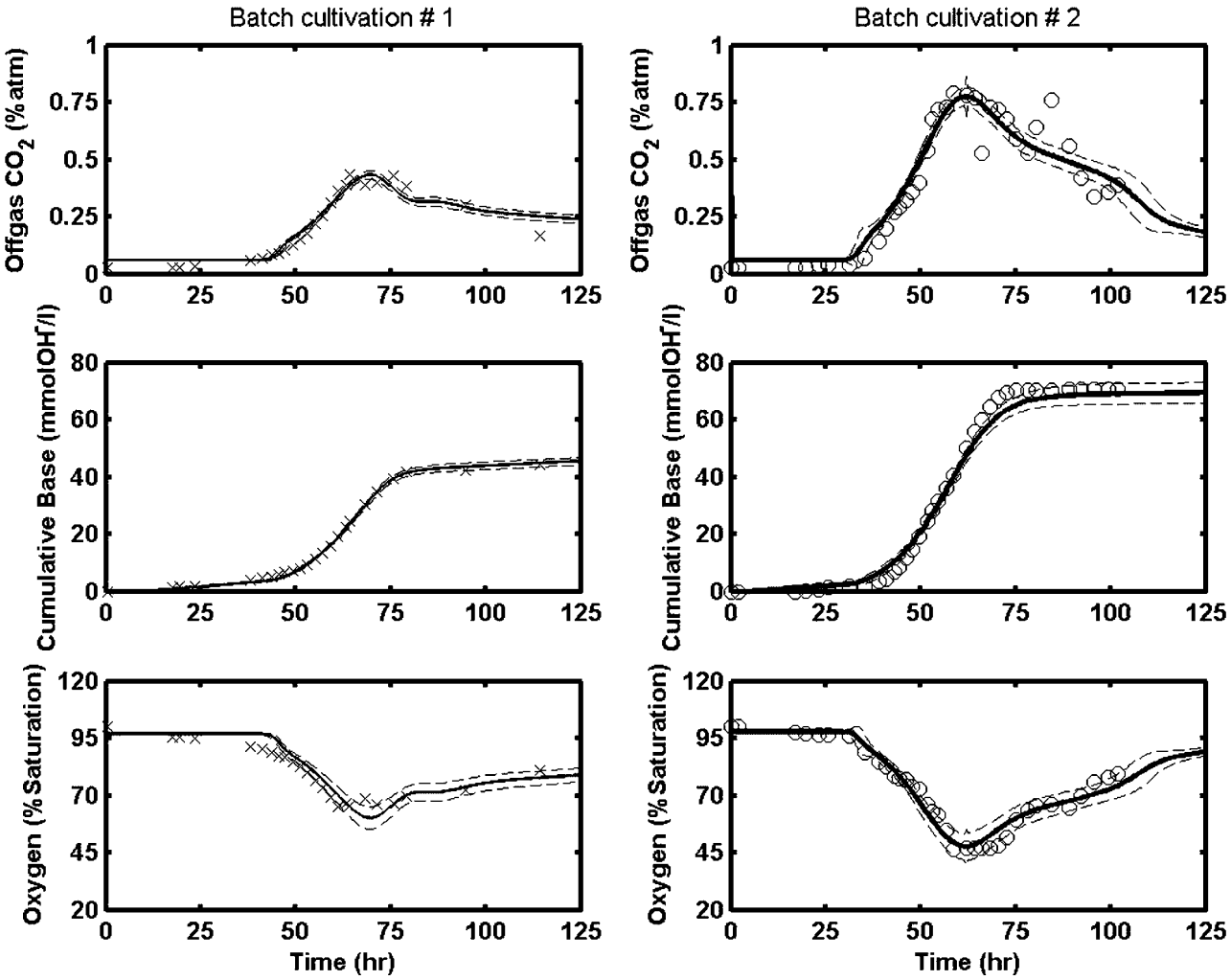


Figure 2. Model fits to the antibiotic production together with 95% upper and lower confidence bounds: (×) data batch 1, (–) model batch 1, (○) data batch 2, and (—) model batch 2.

Typical to batch fermentations, biomass growth and associated substrate uptake started following a lag phase. The duration of the lag phase was different in the two batches, which was reflected in different estimates of the parameter t_{lag} (Table VI). The onset of the exponential growth phase corresponded to the start of the uptake of ammonium and glucose. One observes from Figure 1 that the model was able to capture sufficiently well the measured trends in the biomass, glucose and ammonium concentrations for two batches. However, there was a significant discrepancy between the model fit and the measured phosphate, which was also confirmed by the relatively high values of the MSE criterion for phosphate (see Table VII, where the MSE criterion of phosphate is the highest of all). Overall, this indicates that the linear relations assumed in the model hold for the relation between biomass growth kinetics and the glucose and ammonium uptake kinetics, but are not entirely valid for the phosphate uptake kinetics. When comparing the model and the data, it appeared that the onset of phosphate uptake preceded the onset of the biomass growth. In other words, a gap in our process knowledge was detected. The phosphate that was taken up may be stored internally as polyphosphate before being used for biomass growth, which would explain why phosphate uptake preceded biomass growth. This aspect of the model can be further improved by defining an additional process in the model to include polyphosphate storage.

The Antibiotic Production

Antibiotic production was not associated with growth, but started following the cessation of the exponential growth phase, a situation that occurred when the phosphate became limiting. This behavior is what is normally observed for actinorhodin production by *S. coelicolor* (see, e.g., Bystrykh et al. 1996; Doull and Vining, 1990; Hobbs et al. 1990). The

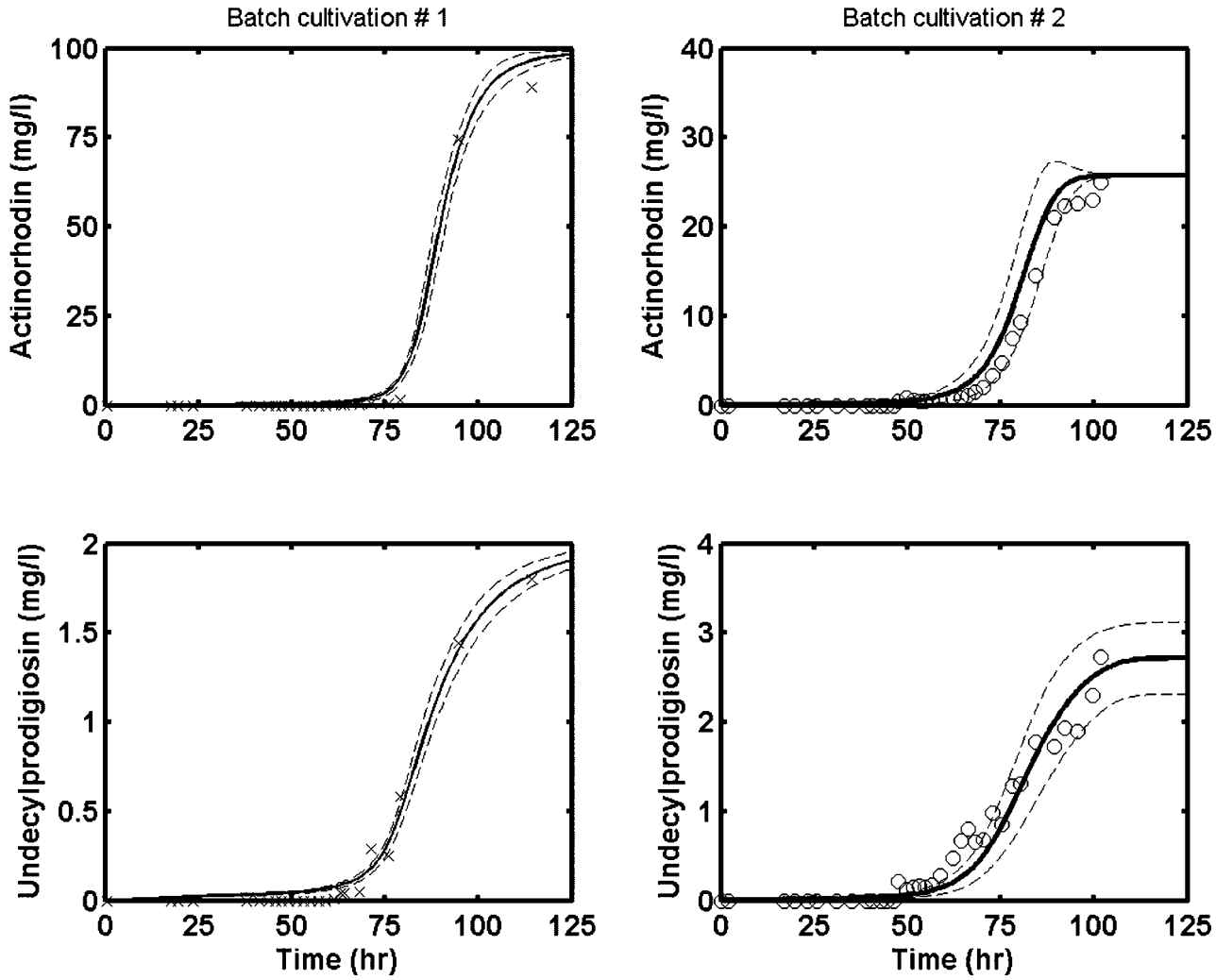


Figure 3. Model fits to the off-gas and base addition data together with 95% upper and lower confidence interval: (×) data batch 1, (–) model batch 1, (○) data batch 2, and (—) model batch 2.

actinorhodin production reached around 100 mg/L in batch 1, which is 3–4 times the amount obtained in batch 2 (see Fig. 2). Faster growth was observed in batch 2 compared to batch 1, which may influence the production. Unlike batch 1, most of the glucose in batch 2 was consumed for the biomass growth. Concerning the RED production, a similar level of production was observed in both batches.

The antibiotic production patterns were nicely captured by the model, which was also supported by the low values of MSE for both antibiotic products (Table VII). Hence, the model fits that were obtained using a modified Luedeking–Piret equation including the logistic expression was found adequate for the given experimental data. Furthermore, the model fits also illustrate that the assumption that the onset of secondary metabolism for antibiotic production is due to phosphate limitation (modeled by an inverted Monod expression) is justified. Moreover, the measured antibiotic production showed that the RED production preceded slightly the onset of actinorhodin production. This observation was included in the model by setting the phosphate threshold in the phosphate limitation function to 0.01 mmol P/L and 0.1 mmol P/L for actinorhodin and RED production, respectively.

The Dissolved Oxygen, Off-Gas CO_2 and Cumulative Base Addition

The model fits of the off-gas CO_2, dissolved oxygen and cumulative base addition data are shown in Figure 3 together with their corresponding 95% uncertainty bounds. The cumulative base addition data were obtained from the model by including a simple and ideal pH controller, which compensates any changes made to pH by adding either base or acid (a detailed discussion of mixed weak acid/base equilibrium reactions particularly the bicarbonate buffer versus CO_2 stripping rate is provided in Supplementary

Table VI. Model identification results using the data from two batch cultivations with *S.coelicolor*.

	Estimates with 95% CI[a]		Correlation matrix (Batch 2)									
θ	Batch 1	Batch 2	Y_{SX}	i_{NX}	i_{PX}	μ_{max}	K_{PO}	β_{ACT}	β_{RED}	t_{lag}	P_{in}	K_La
Y_{SX}	0.50 ± 0.03	0.49 ± 0.02	1	−0.36	−0.21	−0.10	−0.13	−0.27	−0.26	−0.04	−0.51	0.00
i_{NX}	0.24 ± 0.01	0.20 ± 0.01		1	0.63	−0.06	−0.04	0.05	0.07	−0.01	−0.12	0.02
i_{PX}	0.019 ± 0.001	0.01 ± 4.7E-5			1	−0.40	−0.42	−0.28	−0.18	−0.19	−0.30	0.00
μ_{max}	0.13 ± 0.02	0.29 ± 0.25				1	0.99	0.53	0.38	0.81	0.18	0.00
K_{PO}	0.79 ± 0.30	3.05 ± 3.5					1	0.57	0.41	0.75	0.19	0.00
β_{ACT}	3.89E-3 ± 1.1E-3	2.60E-3 ± 1.5E-3						1	0.36	0.26	0.17	0.00
β_{RED}	5.02E-5 ± 7.4E-6	4.70E-5 ± 2.1E-5							1	0.18	0.15	0.00
t_{lag}	45.28 ± 0.9	33.73 ± 2.7								1	0.17	0.00
P_{in}	1.27 ± 0.09	1.17 ± 0.07									1	−0.24
K_La	160.65 ± 16.4	114.08 ± 10.8										1

[a]95% confidence interval of the parameter estimates.

Material 3). One observes that the model describes the measured trends of these physical and chemical data sufficiently well. For these fits, the MSE is found lower albeit the MSE of the off-gas CO_2 is slightly higher, it is deemed acceptable as the dynamic trend was captured nicely. Overall, this confirms the adequacy of the chemical and physical model chosen along with the assumptions made during the model building (see above).

Summary

The integrated model of antibiotic production has in total 60 parameters, of which 38 parameters were obtained from literature sources (see Table V), 12 parameters were obtained from the experimental observations/routine fermentation measurements that were available (see Table V), and the remaining 10 parameters were estimated by model identification (see Table VI).

Concerning the model identification, the accuracy of the model is overall found adequate except for the phosphate data, which is found extremely inaccurate. The latter is hypothesized to be a model structure deficiency (see above). The uncertainty of model fits is also found overall low for the glucose, biomass and ammonium (see Fig. 1), while the uncertainty on the model fits to the antibiotic (particularly for the undecylprodigiosin), off-gas CO_2, base and dissolved oxygen was relatively higher (Figs. 2 and 3). These results indicate that the errors on the estimated parameters are apparently propagated differently to the different model outputs (see Table VI). It is also noteworthy that while the errors on the estimates of μ_{max} and K_{PO} were quite high (see above), the resulting errors on the predictions of the glucose, biomass and ammonium measurements were low. This means that high errors on the estimated parameters may not always result in high errors on the fitted variables. The errors on the fitted variables also depend on the correlation matrix of the parameters. It is thus important to complement a modeling study with an appropriate accuracy and uncertainty analysis as it provides an insight into the statistical reliability of model fits. To improve the accuracy and certainty of the model fits, one needs to improve the underlying hypotheses of the model in an iterative way: dedicated experiments versus model-based interpretation.

Table VII. Statistical evaluation of the model fits to the measurements.

	Mean squared error (MSE[a])	
	Batch 1	Batch 2
Glucose (g/L)	0.23	0.22
Biomass (g dw/L)	0.92	1.63
Ammonium (mM)	0.17	1.44
Phosphate (mM)	31.86	110.69
Actinorhodin (mg/L)	0.43	0.29
Undecylprodigiosin (mg/L)	0.33	0.26
Oxygen (% saturation)	0.25	0.33
Off-gas CO_2 (% atm)	2.62	3.52
Base (mM OH^-)	0.23	0.63

[a]$\mathrm{MSE}(\theta) = \frac{1}{M}\sum_{t=1}^{M}\left(\frac{y_{meas}(t)-y(t,\theta)}{\sigma_t}\right)^2$.

Biomass Maintenance in Batch Fermentations: Pirt Versus Herbert

Two maintenance models, the Pirt model and the Herbert model, are widely used. While the Herbert model is believed to provide a more comprehensive description in batch fermentations, the Pirt model is dominantly used in chemostat studies (Esener et al., 1983). To find out which mechanism that best describes the current *S. coelicolor* batch fermentation data, a model based hypothesis testing was carried out. The integrated model developed above was modified to include only Pirt (Table IIB, process 4), only Herbert (Table IIB, process 5 without C-substrate inhibition) and the combined Herbert and Pirt models (Table IIB, process 4 and process 5 without C-substrate inhibition). These three candidate models were then identified (parameters were estimated) using the data from batch 2. The resulting best model fits are shown in Figure 4 for glucose, biomass, off-gas CO_2 and dissolved oxygen measurements.

Neither the Pirt nor the Herbert model was able to provide a satisfactory description of the observations;

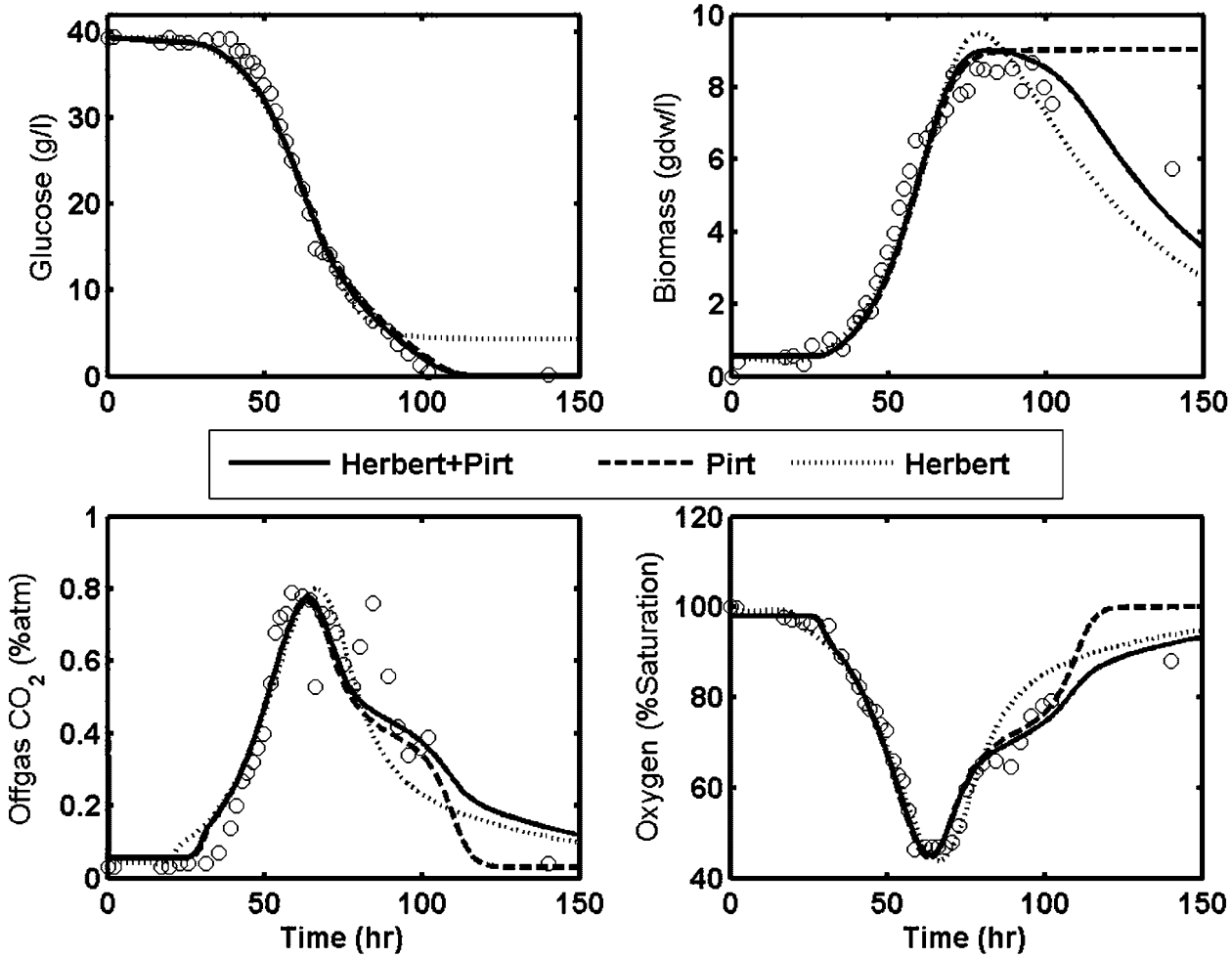


Figure 4. Testing different maintenance hypotheses using the data of batch 2.

particularly the deviations on biomass and glucose were large. While the Pirt model couldn't fit the observed loss of biomass after the stationary phase, the Herbert model underestimated the glucose consumption. The fits of both models could match neither the off-gas CO_2 nor the dissolved oxygen dynamics, particularly during and after the stationary phase. On the other hand, the combined Herbert and Pirt model offered a better fit to both the glucose and biomass observations, and was also better in describing the dynamics of the off-gas CO_2 and the dissolved oxygen. Statistically, the goodness of fit of the combined Herbert and Pirt model was confirmed by the low value of the weighted sum of squared errors (i.e., $J(\boldsymbol{\theta})$ in Eq. 26) criterion (around 360), which was much lower compared to the $J(\boldsymbol{\theta})$ obtained for the Pirt and the Herbert model respectively (408 and 575, respectively). Based on these observations, the combined Herbert and Pirt model was finally selected as the most suitable model structure and included in the model whose fits are shown above (Table V).

The physiological meaning of the combined Herbert and Pirt model could be that biomass uses an external carbon source for the generation of the maintenance energy as long as an external carbon substrate is available (hence the Pirt model). As soon as the external substrate is finished, biomass switches over to endogenous metabolism (oxidizing internal cellular components such as storage carbohydrates) for the maintenance energy, hence Herbert model. Mathematically this is formulated by adding a switch function based on the inverted Monod term for carbon substrate in the Herbert model (Table IIB, process 5). In other words, the Pirt and Herbert models are combined in a complementary way, which is quite different from the complicated approach used by Martens et al. (1995) for the growth of hybridoma cells in steady-state continuous culture. In particular, the proposed new combination allows one to describe a complete growth profile of a microbial culture in batch cultivation. This provides a more comprehensive description of the maintenance process in *S. coelicolor* batch fermentations and should be used in fermentation models.

First-Principles Modeling Within a PAT Framework

Overall, the integrated fermentation model developed in this study using a first-principles approach provided valuable description of the system at a macroscopic level. From a PAT perspective, first-principles modeling supported by modern identification techniques (e.g., confidence level, correlation, uncertainty, etc.) was found to provide a deeper insight into the process. For instance, the model-based data interpretation presented above helped further understand the process and formulating two hypotheses: polyphosphate storage during growth and maintenance mechanisms. Further it also helped identify possible areas of further research to increase the process knowledge and model prediction certainty, for example, antibiotic production kinetics. While the data interpretation could have been performed in a more traditional way, using the model helps ground the data interpretation and process understanding on a systematic basis.

The current model is not an endpoint. With the process knowledge currently incorporated in the unstructured model, the model could be used to further test these hypotheses, hence to further improve the mechanistic understanding of antibiotic production with *S. coelicolor*. Moreover, the model can also be helpful for a number of other purposes including (1) model-based process monitoring, for example, on-line prediction of biomass, glucose, phosphate. To achieve this ambitious purpose, however, model identification remains critical and important, which requires further dedicated research, (2) developing novel control strategies, for example, pH controller, dosage of limiting substrates, feed strategy, etc., and (3) process optimization purposes.

The integrated model developed here is valuable on its own as it can be applied easily to describe other fermentation systems that use different cultures and make different products. The only requirement is model identification to estimate the physiological parameters intrinsic for different cultures as well as operational parameters unique to different fermentations.

The proposed matrix notation and the structured way of building the model illustrated in this paper, that is, developing first a model of the biological processes followed by models of the chemical and physical processes, is helpful in structuring process knowledge available about this system. We believe that the matrix notation based model development is a promising method for presenting and communicating complex *first-principles* models in a simple and efficient way, which is a necessity in a PAT framework.

Conclusions

A matrix notation based model development methodology was successfully evaluated for describing the main biological, chemical (pH) and physical (gas–liquid exchange) processes ongoing during phosphate limited batch fermentations with *S. coelicolor* for actinorhodin and undecylprodigiosin production. The integrated fermentation model is general and can be adopted easily to describe other systems. Overall the matrix notation was found efficient during the model building in this study and is generally expected to facilitate the model development within the PAT initiative.

The model was successfully identified using a data set comprising biological variables, base addition and off-gas data from two batches. While the model described well the glucose and ammonium uptake kinetics, it however failed to describe well the phosphate uptake. The latter is found to precede the onset of the exponential growth phase, and is hypothesized to be stored internally as polyphosphate before it is used in biomass formation. The antibiotic production kinetics was described adequately by using a modified Luedeking–Piret equation with a phosphate inhibition function. The maintenance process was best described using a combined Herbert and Pirt model, which was supported by a model-based hypothesis testing.

Overall the model provided a complete and statistically adequate description of the fermentation system. The model can now be used to further improve the mechanistic understanding within a PAT application, for example, further validate the hypotheses, for model-based process monitoring or for control strategy development.

Supplementary Material Available

Three supplementary materials are available for download. The first one provides details about the estimation of the Herbert decay rate constant using the data from two dedicated batch cultivations, the second one is providing the complete matrix presentation of the integrated fermentation model and the third supplementary material provides a detailed discussion of mixed weak acid/base equilibrium reactions, particularly the bicarbonate buffer versus CO_2 stripping rate.

Nomenclature (Model Components Are as Defined in the Corresponding Tables)

$\mathbf{c}$	composition matrix
$\mathbf{C}$	vector of model components in liquid (mmol/L)
C_G	molar concentration of gas in the reactor (mmol/L)
C_G^{in}	molar concentration of gas in the inflow (mmol/L)
C_G^{out}	molar concentration of gas in the outflow (mmol/L)
$\mathbf{COV}(\boldsymbol{\theta})$	covariance matrix of parameter estimates
$\mathbf{COV}(\mathbf{y_k})$	covariance matrix of model predictions
diag	diagonal elements of matrix
D_{O_2}	diffusion coefficient of O_2 in water at 25°C (cm^2/s)
D_{CO_2}	diffusion coefficient of CO_2 in water at 25°C (cm^2/s)
D_{N_2}	diffusion coefficient of N_2 in water at 25°C (cm^2/s)
f_{XI}	inert particulate content of biomass (g/g DW)
i_{NX}	nitrogen content of biomass (N-mmol /C-mmol X)
i_{PX}	phosphorus content of biomass (P-mmol /C-mmol X)
i_{NRED}	nitrogen content of RED (N-mmol/C-mmol RED)

$J(\boldsymbol{\theta})$	weighted sum of squared errors
$k_{f,NH}$	apparent forward rate constant for NH_4 dissociation (s^{-1})
k_{f,CO_2}	apparent forward rate constant for CO_2 dissociation (s^{-1})
k_{f,H_2PO_4}	apparent forward rate constant for H_2PO_4 dissociation (s^{-1})
$k_{f,W}$	apparent forward rate constant for water dissociation (s^{-1})
k_d	biomass decay coefficient (h^{-1})
K_{IP}	threshold phosphate concentration (P-mmol/L)
K_{PO}	Monod half-saturation coefficient for phosphate (P-mmol/L)
K_S	Monod half-saturation coefficient for substrate (C-mmol/L)
K_O	Monod half-saturation coefficient for oxygen (O-mmol/L)
K_{NH_3}	Monod half-saturation coefficient for ammonia (N-mmol/L)
K_{H_2O}	Henry's gas–liquid equilibrium constant for O_2 (mmol/(L atm))
K_{HCO_2}	Henry's gas–liquid equilibrium constant for CO_2 (mmol/(L atm))
K_{HN_2}	Henry's gas–liquid equilibrium constant for N_2 (mmol/(L atm))
$K_La_{O_2}$	mass transfer rate coefficient for ($O_2\,h^{-1}$)
$K_La_{CO_2}$	mass transfer rate coefficient for ($CO_2\,h^{-1}$)
$K_La_{N_2}$	mass transfer rate coefficient for ($N_2\,h^{-1}$)
$K_La_{NH_3}$	mass transfer rate coefficient for ($NH_3\,h^{-1}$)
m_S	maintenance coefficient (C-mmol S/(C-mmol X h))
M	total number of measured variables used in parameter estimation
ME	mean error
MSE	mean squared error
MW_X	molecular weight of biomass (mg/C-mmol X)
MW_S	molecular weight of substrate (mg/C-mmol S)
MW_{act}	molecular weight of ACT (g/C-mmol ACT)
MW_{red}	molecular weight of RED (g/C-mmol RED)
n	power coefficient in the empirical relationship of mass transfer coefficients
n_G	total number of moles in gas phase (mmol)
n_G^{in}	total number of moles in inflow gas (flux) (mmol/day)
n_G^{out}	total number of moles in outflow gas (flux) (mmol/day)
$n_G^{exchange}$	total number of moles exchanged between gas and liquid phases (mmol/day)
N	total number of measurements for each variable
p	index of gaseous components in the model
PAT	process analytical technology
pH	negative logarithm of hydrogen ions
pK_{NH}	apparent equilibrium constant for NH_4 dissociation
pK_{HCO_3}	apparent equilibrium constant for CO_2 dissociation
$pK_{H_2PO_4}$	apparent equilibrium constant for H_2PO_4 dissociation
pK_w	apparent equilibrium constant for water dissociation
P_G^{in}	total pressure of gas in the inflow (atm)
P_G^{out}	total pressure of gas in the outflow (off-gas) (atm)
P_{O_2}	partial pressure of O_2 in dry air or gas (atm)
P_{CO_2}	partial pressure of CO_2 in dry air or gas (atm)
P_{NH_3}	partial pressure of NH_3 in dry air or gas (atm)
P_{N_2}	partial pressure of N_2 in dry air or gas (atm)
$\mathbf{r}$	overall conversion rate vector for each component (mmol/L day)
R	universal gas constant (L atm/(K mol))
R_{ij}	correlation between the ith and the jth parameters
Q	flow rate (liquid) (L/h)
Q_G^{in}	inflow rate of gas (L/h)
Q_G^{out}	outflow rate of gas (off-gas) (L/h)
$S_{O_2}^*$	saturation concentration of O_2 in liquid (mmol/L)
$S_{CO_2}^*$	saturation concentration of CO_2 in liquid (mmol/L)
$S_{N_2}^*$	saturation concentration of N_2 in liquid (mmol/L)
$S_{NH_3}^*$	saturation concentration of NH_3 in liquid (mmol/L)
S_{ACT}^{max}	maximum level of ACT production (C-mmol/L)
S_{RED}^{max}	maximum level of RED production (C-mmol RED/L)
$\mathbf{S}$	stoichiometric matrix
t_{lag}	lag time prior to onset of biomass growth (h^{-1})
T_G^{in}	temperature of gas in the inflow (K)
T_G^{out}	temperature of gas in the outflow (off-gas) (K)
T	temperature of the reactor (K)
u_1	stoichiometric coefficient of oxygen in maintenance metabolism (O-mmol/C-mmol)
V_G	volume of gas in the reactor (L)
V_L	liquid volume in the reactor (L)
w_1	stoichiometric coefficient of oxygen in decay metabolism (O-mmol/C-mmol)
w_2	stoichiometric coefficient of nitrogen in decay metabolism (O-mmol/C-mmol)
w_3	stoichiometric coefficient of phosphate in decay metabolism (O-mmol/C-mmol)
y_1	stoichiometric coefficient of oxygen in product metabolism (O-mmol/C-mmol)
y_2	stoichiometric coefficient of nitrogen in product metabolism (O-mmol/C-mmol)
y_3	stoichiometric coefficient of CO_2 in product metabolism (O-mmol/C-mmol)
y_4	stoichiometric coefficient of H^+ in product metabolism (O-mmol/C-mmol)
$y_{k,meas}$	measurements of the kth variable
y_k	model prediction of the kth variable
$y_{k,1-\alpha}$	confidence interval of model predictions at $1-\alpha$ confidence level
$\frac{\partial y_k}{\partial \theta}$	sensitivity of the measured variables, $\mathbf{y_k}$ to the parameter $\boldsymbol{\theta}$
Y_{SACT}	actinorhodin (ACT) yield on substrate (C-mmol Act/C-mmol S)
Y_{SRED}	undecylprodigiosin (RED) yield on substrate (C-mmol RED/C-mmol S)
Y_{XO}	biomass yield on oxygen (C-mmol X/C-mmol O)
Y_{SX}	biomass yield on substrate (C-mmol X/C-mmol S)
z_1	stoichiometric coefficient of oxygen in growth metabolism (O-mmol/C-mmol)
z_2	stoichiometric coefficient of nitrogen in growth metabolism (O-mmol/C-mmol)
z_3	stoichiometric coefficient of phosphate in growth metabolism (O-mmol/C-mmol)
z_4	stoichiometric coefficient of CO_2 in growth metabolism (O-mmol/C-mmol)
z_5	stoichiometric coefficient of H^+ in growth metabolism (O-mmol/C-mmol)

Greek Letters

α	significance level at which confidence of parameter and prediction errors are estimated, whereas $(1-\alpha)$ means confidence level
α_{ACT}	growth associated ACT production constant (C-mmol Act/C-mmol X)
α_{RED}	growth associated RED production constant (C-mmol RED/C-mmol X)

β_{RED}	non-growth associated RED production rate (C-mmol RED/(C-mmol X h)
β_{ACT}	non-growth associated ACT production rate (C-mmol Act/(C-mmol X h)
γ_{RED}	degree of reduction of RED (mmol e^-/C-mmol RED)
γ_{ACT}	degree of reduction of ACT (mmol e^-/C-mmol ACT)
γ_O	degree of reduction of oxygen (mmol e^-/C-mmol S)
γ_S	degree of reduction of substrate (mmol e^-/C-mmol S)
γ_X	degree of reduction of biomass (mmol e^-/C-mmol X)
μ_{max}	maximum growth rate of biomass (h^{-1})
$\boldsymbol{\rho}$	process rate vector (mmol/L day)
$\sigma_{k,t}$	standard error of measurements for the kth variable at time instant t
$\boldsymbol{\theta}$	estimated parameters subset of the model
$\boldsymbol{\theta}_{1-\alpha}$	confidence interval of estimated parameters at $1-\alpha$ confidence level

The research work of Dr. Gürkan Sin is financed by a H.C. Ørsted postdoctoral fellowship of the Technical University of Denmark. The PhD projects of Nanna Petersen and Peter Ödman are supported by a grant from the Innovative Bioprocess Technology Research Consortium financed by the Danish Research Council for Technology and Production Sciences, Chr. Hansen A/S, Danisco A/S and Novozymes A/S. The constructive comments of the anonymous reviewers of the manuscript are gratefully acknowledged.

References

Baltes M, Schneider R, Sturm C, Reuss M. 1994. Optimal experimental design for parameter estimation in unstructured growth models. Biotechnol Prog 10:480–488.

Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, et al. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417(6885):141–147.

Bruheim P, Butle rM, Ellingsen TE. 2002. A theoretical analysis of the biosynthesis of actinorhodin in a hyper-producing *Streptomyces lividans* strain cultivated on various carbon sources. Appl Microbiol Biotechnol 58:735–742.

Bystrykh LV, FernandezMoreno MA, Herrema JK, Malpartida F, Hopwood DA, Dijkhuizen L. 1996. Production of actinorhodin-related "blue pigments" by *Streptomyces coelicolor* A3(2). J Bacteriol 178:2238–2244.

CRC handbook of chemistry and physics. 2006–2007. 87th Edition (on-line at http://www.hbcpnetbase.com).

Daae E, Ison A. 1998. A simple structured model describing the growth of *Streptomyces lividans*. Biotechnol Bioeng 58:263–266.

Dassau E, Zadok I, Lewin DR. 2006. Combining six-sigma with integrated design and control for yield enhancement in bioprocessing. Ind Eng Chem Res 45:8299–8309.

Doull JL, Vining LC. 1990. Nutritional control of actinorhodin production by *Streptomyces coelicolor* A3(2): Suppressive effect of nitrogen and phosphate. Appl Microbiol Biotechnol 32:449–454.

Esener AA, Roels J, Kossen NWF. 1983. Theory and applications of unstructured growth models: Kinetic and energetic aspects. Biotechnol Bioeng 25:2803–2841.

FDA. 2004. PAT—A framework for innovative pharmaceutical development, manufacturing and quality assurance. Guidance for Industry (http://www.fda.gov/cder/guidance/index.htm).

Gujer W, Larsen TA. 1995. The implementation of biokinetics and conservation principles in ASIM. Water Sci Technol 31(2):257–266.

Heijnen JJ. 1999. Bioenergetics of microbial growth. In: Flickinger MC, Drew SW, editors. Encyclopedia of bioprocess technology: Fermentation, biocatalysis and bioseparation. New York: John Wiley & Sons Inc. p 267–291.

Henze M, Gujer W, Mino T, van Loosdrecht MCM. 2000. Activated sludge models ASM1, ASM2, ASM2d and ASM3. IWA Scientific and Technical Report No. 9. London, UK: IWA Publishing.

Hobbs G, Frazer CM, Gardner DCJ, Flett F, Oliver SG. 1990. Pigmented antibiotic production by *Streptomyces coelicolor* A3(2): Kinetics and the influence of nutrients. J Gen Microbiol 136:2291–2296.

Holmberg A. 1982. On the practical identifiability of microbial growth models incorporating Michealis-Menten type nonlinearities. Math Biosci 62:23–43.

Junker BH, Wang HY. 2006. Bioprocess monitoring and computer control: key roots of the current PAT initiative. Biotechnol Bioeng 95:226–261.

Kieser T, Bibb M, Buttner M, Chater K, Hopwood DA. 2000. Practical *Streptomyces* genetics. The John Innes Foundation.

King R. 1997. A structured mathematical model for a class of organisms. I. Development of a model for *Streptomyces tendae* and application of model-based control. J Biotechnol 52:219–234.

Kiviharju K, Salonen K, Leisola M, Eerikainen T. 2006. Modeling and simulation of *Streptomyces peucetius var. caesius* N47 cultivation and ε-rhodomycinone production with kinetic equations and neural networks. J Biotechnol 126:365–373.

Maes I, Van Liedekerke B. 2006. The need for a broader perspective if process analytical technology implementation is to be successful in the pharmaceutical sector. J Pharm Innovation 1:19–21.

Martens DE, Sipkerna EM, de Gooijer CD, Beuvery EC, Tramper J. 1995. A combined cell-cycle and metabolic model for the growth of hybridoma cells in steady-state continuous culture. Biotechnol Bioeng 48:49–65.

Martín JF. 2004. Phosphate control of the biosynthesis of antibiotics and other secondary metabolites is mediated by the PhoR-PhoP system: An unfinished story. J Bacteriol 186:5197–5201.

Munack A. 1991. Optimisation of sampling. In: Rehm , Reed , editors. Biotechnology, Vol. 4. Weinheim: Verlag Chemie. p 252–264.

Mundry C, Kuhn K-P. 1991. Modelling and parameter identification for batch fermentations with *Streptomyces tendae* under phosphate limitation. Appl Microbiol Biotechnol 35:306–311.

Musvoto EV, Wentzel MC, Loewenthal RE, Ekama GA. 1997. Kinetic-based model for mixed weak acid/base systems. Water SA 23:311–322.

Nielsen J, Villadsen J. 1992. Modelling of microbial kinetics. Chem Eng Sci 47:4225–4270.

NIST Chemistry WebBook. 2005. NIST Standard Reference Database Number 69 (http://webbook.nist.gov/chemistry/).

Noorman HJ, Heijnen JJ, Luyben KChAM. 1991. Linear relations in microbial reaction systems: A general overview of their origin, form and use. Biotechnol Bioeng 38:603–618.

Omlin M, Reichert P. 1999. A comparison of techniques for the estimation of model prediction uncertainty. Ecol Model 115:45–59.

Pirt SJ. 1965. Maintenance energy of bacteria in growing cultures. Proc R Soc London Series B Biol Sci 16:224.

Power M. 1993. The predictive validation of ecological and environmental models. Ecol Model 68:33–350.

Redenbach M, Kieser HM, Denapaite D, Eichner A, Cullum J, Kinashi H, Hopwood DA. 1996. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. Mol Microbiol 21(1):77–96.

Roels JA. 1980. Application of macroscopic principles to microbial metabolism. Biotechnol Bioeng 22:2457–2514.

Royce PN. 1992. Effect of changes in the pH and carbon dioxide evolution rate on the measured respirometry quotient of fermentation. Biotechnol Bioeng 40:1129–1138.

San K-Y, Stephanopoulos G. 1984. Studies on on-line bioreactor identification. II. Numerical and experimental results. Biotechnol Bioeng 26:1189–1197.

Seber G, Wild C. 1989. Nonlinear regression. New York: Wiley.

Siano SA. 1995. On the use of pH control reagent addition rate for fermentation monitoring. Biotechnol Bioeng 47:651–665.

Sötemann SW, Musvoto EV, Wentzel MC, Ekama GA. 2005. Integrated biological, chemical and physical processes kinetic modelling Part 1 -

Anoxic-aerobic C and N removal in the activated sludge system. Water SA 31(4):529–544.
Teusink B, Smid EJ. 2006. Modelling strategies for the industrial exploitation of lactic acid bacteria. Nat Rev Microbiol 4(1):46–56.
Thilakavathi M, Basak T, Panda T. 2007. Modeling of enzyme production kinetics. Appl Microbiol Biotechnol 73:991–1007.
Verduyn C, Postma E, Scheffers WA, van Dijken JP. 1992. Effect of benzoic acid on metabolic fluxes in yeast: A continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 8:501–517.
Volcke EIP. 2006. Modelling, analysis and control of partial nitritation in a SHARON reactor. PhD thesis, Ghent University, Belgium, p. 300.

www.bioeng.kt.dtu.dk

Department of Chemical and Biochemical Engineering
Center for Process Engineering and Technology (PROCESS)
Technical University of Denmark
Søltofts Plads
Building 229
DK-2800 Kgs. Lyngby
Denmark
Tel: (+45) 45 25 28 00
Fax: (+45) 45 93 29 06
Email: kt@kt.dtu.dk